

# Testing for Common Breaks in a Multiple Equations System\*

Tatsushi Oka<sup>†</sup>

Pierre Perron<sup>‡</sup>

First version: April 19, 2009

This version: May 27, 2016

## Abstract

The issue addressed in this paper is that of testing for common breaks across or within equations of a multivariate system. Our framework is very general and allows integrated regressors and trends as well as stationary regressors. The null hypothesis is that breaks in different parameters (either regression coefficients or elements of the covariance matrix of the errors) occur at a common locations or are separated by some positive fraction of the sample size. Under the alternative hypothesis, the break dates are not the same and also need not be separated by a positive fraction of the sample size across parameters. The test considered is the quasi-likelihood ratio test assuming normal errors, though as usual the limit distribution of the test remains valid with non-normal errors. Also of independent interest, we provide results about the rate of convergence when searching over all possible partitions subject only to the requirement that each regime of different parameters contains at least as many observations as some positive fraction of the sample size. Simulation results show that the test has good finite sample properties. We also provide an application to various measures of inflation to illustrate its usefulness.

**Keywords:** change-point, segmented regressions, break dates, hypothesis testing, a system of multiple equations.

---

\*Perron's work is supported by the National Science Foundation under Grant (SES-0649350). Oka's work is supported by Singapore Ministry of Education Academic Research Fund (FY2015-FRC3-003).

<sup>†</sup>Department of Economics, National University of Singapore, Singapore (oka@nus.edu.sg).

<sup>‡</sup>Department of Economics, Boston University, USA (perron@bu.edu).

# 1 Introduction

Issues related to structural change have been extensively studied in the statistics and econometrics literature (see Csörgö and Horváth, 1997; Perron et al., 2006, for a comprehensive review). In the last twenty years or so, substantial advances have been made in the econometrics literature to cover models at a level of generality that makes them relevant across time-series applications in the context of unknown change points. For example, Bai (1994, 1997) studies the least squares estimation of a single change point in regressions involving stationary and/or trending regressors. Bai and Perron (1998, 2003) extend the testing and estimation analysis to the case of multiple structural changes and present an efficient algorithm. Hansen (1992) and Kejriwal and Perron (2008) consider regressions with integrated variables. Andrews (1993) and Hall and Sen (1999) consider nonlinear models estimated by generalized method of moments. Bai (1995, 1998) studies structural changes in least absolute deviation regression, while Qu (2008), Su and Xiao (2008) and Oka and Qu (2011) analyze structural changes in quantile regression. Hall, Han, and Boldea (2012) and Perron and Yamamoto (2014) consider structural changes in linear models with endogenous regressors. The studies on structural changes in panel data model include Bai (2010), Kim (2011), Baltagi, Feng, and Kao (2015) and Qian and Su (2016) for linear panel data models and Breitung and Eickmeier (2011), Cheng, Liao, and Schorfheide (2016), Corradi and Swanson (2014), Han and Inoue (2015) and Yamamoto and Tanaka (2015) for factor models.

The literature on structural breaks in a multiple equations system includes Bai, Lumsdaine, and Stock (1998), Bai (2000) and Qu and Perron (2007) among others. Their analysis relies on the assumption of common breaks, under which breaks in different basic parameters (either regression coefficients or elements of the covariance matrix of the errors) occur at a common location or are separated by some positive fraction of the sample size (i.e., asymptotically distinct).<sup>1</sup> Bai et al. (1998) assume a single common break across equations of a multivariate system with stationary regressors and trends as well as cointegrated systems. For the case of multiple common breaks, Bai (2000) analyzes vector autoregressive models for stationary variables and Qu and Perron (2007) cover multiple system equations, allowing for more general stationary regressors and arbitrary restrictions across the parameters. Under the framework of Qu and Perron (2007), Kurozumi and Tuvaandorj (2011) propose model selection procedures for a system of equations with multiple common breaks and Eo and Morley (2015) consider a confidence set for the common break date based on inverting

---

<sup>1</sup> Note that the concept of common breaks analyzed here is quite distinct from the notion of co-breaking or co-trending (e.g., Hatanaka and Yamada, 2003; Hendry and Mizon, 1998). In this literature, the focus is to assess whether some linear combination of series with breaks do not have a break, a concept akin to that of cointegration.

the likelihood ratio test. In the literature on common breaks, it is documented that common breaks allow more precise estimates of the break date in a multivariate system. Given unknown break dates, however, an issue of interest for most applications concerns the validity of the assumption of common breaks.<sup>2</sup> To our knowledge, no test has been proposed to address this issue.

Our paper addresses three outstanding issues with regard to the test for common breaks. First, we propose the quasi-likelihood ratio test for common breaks under a very general framework. We consider a system of multiple equations under the likelihood framework with normal errors, though as usual the limit distribution of the proposed test remains valid with non-normal, serially dependent and heteroskedastic errors.<sup>3</sup> Our framework allows integrated regressors and trends as well as stationary regressors as the one in Bai et al. (1998) and also can be applied for models with multiple breaks and arbitrary restrictions across the parameters as the one in Qu and Perron (2007). Thus, our results can apply for a system of multiple equations considered in the existing studies.

Second, we propose the test for common breaks not only across equations within a multivariate system, but also within an equation. As in Bai et al. (1998), the issue of common breaks is often associated with breaks occurring across equations within a system, whereas one may want to test common breaks in the parameters within a regression equation, whether a single equation or a system of multiple equations are considered. More precisely, the null hypothesis of interest is that some subsets of the basic parameters share one or more common break dates, so that each regime is separated by some positive fraction of the sample size. Under the alternative hypothesis, the break dates are not the same and also need not be separated by a positive fraction of the sample size, or be asymptotically distinct.

Third, we derive asymptotic properties of the quasi-likelihood and its estimators, allowing for the possibility that break dates of different basic parameters may not be asymptotically distinct. This poses an additional layer of difficulty, since the existing studies establish the consistency and rate of convergence of estimators only when break dates are assumed to be either common location or asymptotically distinct, at least under the level of generality adopted here. Moreover, we establish the results in the presence of integrated regressors and trends as well as stationary regressors. This is by itself a noteworthy contribution. This asymptotic result will allow us to derive the limit distribution of our test statistic under the null hypothesis and also facilitate asymptotic power analysis under fixed and local

---

<sup>2</sup> The common breaks assumption is also used in the literature on panel data (see Bai, 2010; Kim, 2011; Baltagi et al., 2015, for example). In this paper, we consider the multiple equations system in which the number of equations are finite, and thus panel data models are out of our scope.

<sup>3</sup> For a discussion on testing under possible dynamic misspecification, see Corradi and Swanson (2006), who consider specification tests.

alternatives. We can show that our test is consistent under the fixed alternative and also has non-trivial local power.

There is one additional layer of difficulty compared to Bai and Perron (1998) or Qu and Perron (2007). In their analysis, it is possible to transform the limit distribution so that the limit distribution can be evaluated using a closed form solution and thus critical values can be tabulated. Here, no such solution is available and we need to obtain critical values for each case through simulation. This involves simulating the Wiener processes with consistent parameter estimates and evaluating each realization of the limit distribution with and without the restriction of common breaks. While it is conceptually straightforward, the procedure needs to be repeated many times to obtain relevant quantities and can be quite computationally intensive. This is because we need to search over many possible combinations of all the permutations of the break locations for each replication of the simulation. The procedure suggested is nevertheless quick enough to be feasible for common applications, but the computational burden will have an impact on the extent to which we will evaluate the finite sample performance of the test via Monte Carlo simulations. Because of this computational burden, we conduct Monte Carlo simulations only in limited cases, while our simulation result suggests that the test proposed has reasonably good size and power performance even in small samples. Also, we apply our test for common breaks in inflation series, following the work of Clark (2006) to illustrate its usefulness.

The remainder of the paper is as follows. Section 2 introduces models with and without common breaks assumption and describes estimation methods under the quasi-likelihood framework. Section 3 presents assumptions and asymptotic results including the asymptotic null distribution and asymptotic power analysis. Section 4 reports results of some Monte Carlo simulations to evaluate the finite sample properties of our procedures. Section 5 gives an empirical application and Section 6 concludes. Appendix contains all the proofs.

## 2 Models and quasi-likelihood method

In this section, we first introduce models for a multiple equations system under common breaks assumption and then explain the case without common breaks assumption. Subsequently, we describe the quasi-likelihood estimation method for the models assuming normal errors and then propose the quasi-likelihood ratio test for common breaks. For illustration purpose, we also discuss some examples.

As a matter of notation, “ $\xrightarrow{p}$ ” denotes convergence in probability, “ $\xrightarrow{d}$ ” convergence in distribution and “ $\Rightarrow$ ” weak convergence in the space  $D[0, \infty)$  under the Skorohod metric. We use  $\mathbb{R}$ ,  $\mathbb{Z}$  and  $\mathbb{N}$  to denote the set of all real numbers, all integers and all positive integers,

respectively. For a vector  $x$ , we use  $\|\cdot\|$  to denote the Euclidean norm (i.e.,  $\|x\| = \sqrt{x'x}$ ), while for a matrix  $A$ , we use the vector-induced norm (i.e.,  $\|A\| = \sup_{x \neq 0} \|Ax\|/\|x\|$ ). Also, define the  $L_r$ -norm of a random matrix  $X$  as  $\|X\|_r = (\sum_i \sum_j E |X_{ij}|^r)^{1/r}$  for  $r \geq 1$ . As usual,  $a \wedge b = \min\{a, b\}$  and  $a \vee b = \max\{a, b\}$  for any  $a, b \in \mathbb{R}$ . Let  $\circ$  denote the Hadamard product (entry-wise product) and define  $e_i$  as a unit vector having 1 at the  $i^{th}$  entry but 0 at the rest of entries. We use the operator  $\text{vec}(\cdot)$  to convert a matrix into a column vector by stacking the columns of the matrix and the operator  $\text{tr}(\cdot)$  to denote the trace of a matrix. Let  $\mathbb{1}_{\{\cdot\}}$  be the indicator function taking the value one when its argument is true, and zero otherwise. We denote  $[a]$  as the largest integer not greater than  $a$  for  $a \in \mathbb{R}$ . Let  $\text{sgn}(\cdot)$  be the sign function; for  $a \in \mathbb{R}$ ,  $\text{sgn}(a)$  takes 1, 0 and -1 for  $a > 0$ ,  $a = 0$  and  $a < 0$ , respectively.

## 2.1 The models with and without common breaks assumption

Let the data consist of observations  $\{(y_t, x_{tT})\}_{t=1}^T$ , where  $y_t$  is an  $n \times 1$  vector of dependent variables and  $x_{tT}$  is a  $q \times 1$  vector of explanatory variables for  $n, q \in \mathbb{N}$  with a subscript  $t$  indexing a temporal observation and  $T$  denoting the sample size. We allow the set of regressors  $x_{tT}$  to include stationary variables, time trends and integrated processes, while scaling by the sample size  $T$  so that the order of all components is the same. In what follows, we consider

$$x_{tT} = (z'_t, \varphi(t/T)', T^{-1/2}w'_t)'$$

Here,  $z_t$ ,  $\varphi(t/T)$  and  $w_t$  denote a  $q_z \times 1$  vector of stationary regressors, a  $q_\varphi \times 1$  vector of trending variables and a  $q_w \times 1$  vector of integrated regressors, respectively, so that  $q \equiv q_z + q_\varphi + q_w$ , and satisfy

$$\varphi(t/T) := [(t/T), (t/T)^2, \dots, (t/T)^{q_\varphi}]' \quad \text{and} \quad w_t = w_{t-1} + u_{wt},$$

where  $w_0$  is assumed, for simplicity, to be either  $O_p(1)$  random variables or fixed finite constants, and  $u_{wt}$  is a vector of unobserved random variables with zero means. By labelling the regressors  $z_t$  as  $I(0)$ , we mean that the partial sums of the associated noise components satisfy a functional central limit theorem, while we label a variable as  $I(1)$  if it is the accumulation of an  $I(0)$  process. We discuss more details of the conditions in Section 3.

We first explain the case of common breaks through the model in which all of the coefficients and the covariance matrix of the errors change, i.e., a pure structural change model. The model of our interest is a multiple equations system with  $n$  equations and  $T$  time periods, excluding the initial conditions if lagged dependent variables are used as regressors. We denote break dates in the system by  $T_1, \dots, T_m$  with  $m$  denoting the total number of

structural changes and we use the convention that  $T_0 = 0$  and  $T_{m+1} = T$ .<sup>4</sup> The model with a subscript  $j$  indexing a regime for  $j = 1, \dots, m+1$  is given by

$$y_t = (x'_{tT} \otimes I_n) S \beta_j + u_t, \quad (1)$$

for  $T_{j-1} + 1 \leq t \leq T_j$ , where  $I_n$  is an  $n \times n$  identity matrix,  $S$  is an  $nq \times p$  selection matrix with full column rank,  $\beta_j$  is a  $p \times 1$  vector of unknown coefficients, and  $u_t$  is an  $n \times 1$  vector of errors having zero means and covariance matrix  $\Sigma_j$ . The selection matrix  $S$  is usually a matrix involving elements that are 0 or 1 and, hence, specifies which regressors appear in each equation, although it in principle is allowed to have entries that are arbitrary constants. To ease notation, define the  $n \times p$  matrix  $X_{tT} := S'(x_{tT} \otimes I_n)$  so that (1) becomes

$$y_t = X'_{tT} \beta_j + u_t, \quad (2)$$

for  $T_{j-1} + 1 \leq t \leq T_j$  with  $j = 1, \dots, m+1$ .

The set of basic parameters in the  $j^{th}$  regime consists of the coefficients  $\beta_j$  and the covariance matrix  $\Sigma_j$ , and we denote it by  $\theta_j := (\beta_j, \Sigma_j)$  for each regime  $j = 1, \dots, m+1$ . We use  $\Theta_j \subset \mathbb{R}^p \times \mathbb{R}^{n \times n}$  to denote a parameter space for  $\theta_j$  and we also define a product space  $\Theta := \Theta_1 \times \dots \times \Theta_{m+1}$  for a collection  $\theta := (\theta_1, \dots, \theta_{m+1})$ . In the models of (2), we allow for the imposition of a set of  $r$  restrictions through a function  $R : \Theta \rightarrow \mathbb{R}^r$ , given by

$$R(\theta) = 0. \quad (3)$$

Note that the equation in (3) can impose restrictions within and across both equations and regimes. Thus the model in (2) with some restrictions of (3) can accommodate structural break models other than a pure structural change model, such as partial structural change models in which a part of basic parameters are constant across regimes. For a discussion of how general the framework is, see Qu and Perron (2007).<sup>5</sup>

Next, we consider a pure structural change model allowing for the possibility that break dates are not necessarily common across basic parameters. In the equations system with the  $p \times 1$  vector of coefficients, we can assign each coefficient an index from 1 to  $p$  and we then group the  $p$  indices into disjoint subsets  $\mathcal{G}_1, \dots, \mathcal{G}_G \subset \{1, \dots, p\}$  with  $G$  standing for

---

<sup>4</sup> In this paper, we focus on models with multiple structural breaks. One can use the existing tests to check whether structural breaks exist and also apply the sequential testing procedure or information criteria to determine the number of structural breaks. Depending on models of interest, those procedures are used for each equation or a system.

<sup>5</sup> Here, our argument concerns the generality of pure structural changes models together with the restrictions in (3). If one wants to test the validity of the restrictions, then one can resort the likelihood ratio test allowing for the possibility that break dates in different basic parameters are not common. As shown in Theorem 2 below, the break date estimates and the estimates of basic parameters are asymptotically independent, and thus one can derive the limit distribution of the likelihood ratio test for the restriction.

the total number of groups, such that coefficients indexed by elements of  $\mathcal{G}_g$  share the same break dates for each group  $g = 1, \dots, G$  and  $\cup_{g=1}^G \mathcal{G}_g = \{1, \dots, p\}$ . Given a collection  $\{\mathcal{G}_g\}_{g=1}^G$ , we define, for  $(g, j) \in \{1, \dots, G\} \times \{1, \dots, m+1\}$ ,

$$\beta_{gj} := \sum_{l \in \mathcal{G}_g} e_l \circ \beta_j. \quad (4)$$

Without loss of generality, we assume that the covariance matrix  $\Sigma_j$  has break dates that are common to those in the last group  $G$ . If none of the regression coefficients change at the same time as the covariance matrix  $\Sigma_j$ , then  $\mathcal{G}_G$  is simply an empty set.<sup>6</sup>

To denote the break date for the regime  $j$  and the group  $g$ , we use  $k_{gj}$  for  $(g, j) \in \{1, \dots, G\} \times \{1, \dots, m\}$  with the convention that  $k_{g0} = 0$  and  $k_{g,m+1} = T$ . Also, define a collection of break dates as

$$\mathcal{K} := \{\mathcal{K}_1, \dots, \mathcal{K}_G\} \text{ with } \mathcal{K}_g := (k_{g1}, \dots, k_{gm}),$$

for  $g = 1, \dots, G$ . The regression model can be expressed as the one depending on time-varying basic parameters according to the collection  $\mathcal{K}$ :

$$y_t = X'_{tT} \beta_{t,\mathcal{K}} + u_t, \quad (5)$$

where  $\beta_{t,\mathcal{K}} := \sum_{g=1}^G \beta_{g,t,\mathcal{K}}$  and  $E[u_t u'_t] = \Sigma_{t,\mathcal{K}}$  with

$$\beta_{g,t,\mathcal{K}} := \beta_{gj} \text{ for } k_{g,j-1} + 1 \leq t \leq k_{gj} \text{ and } \Sigma_{t,\mathcal{K}} := \Sigma_j \text{ for } k_{G,j-1} + 1 \leq t \leq k_{Gj}, \quad (6)$$

for  $(g, j) \in \{1, \dots, G\} \times \{1, \dots, m+1\}$ . We also use  $\theta_{t,\mathcal{K}} := (\beta_{t,\mathcal{K}}, \Sigma_{t,\mathcal{K}})$  to denote time-varying basic parameters depending on the collection of break dates  $\mathcal{K}$ . We can express  $\theta_j = (\beta_j, \Sigma_j)$  in the model (5) because we have  $\beta_j = \sum_{g=1}^G \beta_{gj}$  from definition. Thus the restrictions (3) can be imposed on the system (5) to accommodate more general pattern of structural breaks as in the one with common breaks.

We use a 0 superscript to denote the true values of the parameters in both (2) and (5). Thus, the true basic parameters and break dates in (2) are denoted by  $\{(\beta_j^0, \Sigma_j^0)\}_{j=1}^{m+1}$  and  $\{T_j^0\}_{j=1}^m$ , respectively, with the convention that  $T_0^0 = 0$  and  $T_{m+1}^0 = T$ , whereas the ones in (5) are denoted by  $\{\beta_{1j}^0, \dots, \beta_{Gj}^0, \Sigma_j^0\}_{j=1}^{m+1}$  and  $\mathcal{K}_g^0 := (k_{g1}^0, \dots, k_{gm}^0)$  with that  $k_{g0}^0 = 0$  and  $k_{g,m+1}^0 = T$  for  $g = 1, \dots, G$ , respectively. Also let  $\mathcal{K}^0 := \{\mathcal{K}_1^0, \dots, \mathcal{K}_G^0\}$ . Given a collection  $\mathcal{K}$ , let  $\theta_{t,\mathcal{K}}^0 := (\beta_{t,\mathcal{K}}^0, \Sigma_{t,\mathcal{K}}^0)$  with a 0 superscript to denote the ones depending on true basic parameters  $\theta^0 := (\theta_1^0, \dots, \theta_{m+1}^0)$  with  $\theta_j^0 := (\beta_j^0, \Sigma_j^0)$  for  $j = 1, \dots, m+1$ .

---

<sup>6</sup> The results in this paper can be extended for the case where parameters of the covariance matrix of the error have distinctive break dates, although additional notations are needed. For sake of notational simplicity, we only consider the case where break dates are common within each covariance matrix.

## 2.2 The estimation and test under the quasi-likelihood framework

We consider the quasi-maximum likelihood estimation method with serially uncorrelated Gaussian errors for the model of (5) with restrictions of (3). Given the collection of break dates  $\mathcal{K}$  and basic parameters  $\theta$ , the Gaussian quasi-likelihood function is defined as

$$L_T(\mathcal{K}, \theta) := \prod_{t=1}^T f(y_t | X_{tT}, \theta_{t,\mathcal{K}}),$$

where

$$f(y_t | X_{tT}, \theta_{t,\mathcal{K}}) := \frac{1}{(2\pi)^{n/2} |\Sigma_{t,\mathcal{K}}|^{1/2}} \exp \left( -\frac{1}{2} \|\Sigma_{t,\mathcal{K}}^{-1/2} (y_t - X'_{tT} \beta_{t,\mathcal{K}})\|^2 \right).$$

To obtain maximum likelihood estimators, we impose a restriction on the set of permissible partitions with a trimming parameter  $\nu > 0$  as follows:

$$\Xi_\nu := \left\{ \mathcal{K} : \min_{1 \leq g \leq G} \min_{1 \leq j \leq m+1} (k_{gj} - k_{g,j-1}) \geq T\nu \right\}.$$

This set of permissible partitions ensures that there are enough observations between any break dates within the same group  $\mathcal{K}_g$ , while it accommodates the possibility that the break dates across different groups are not necessarily separated by the positive fraction of the sample size. The quasi-maximum likelihood estimates, which are denoted by  $(\hat{\mathcal{K}}, \hat{\theta})$ , are obtained by maximizing the restricted log-likelihood function as follows:

$$(\hat{\mathcal{K}}, \hat{\theta}) := \arg \max_{(\mathcal{K}, \theta) \in \Xi_\nu \times \Theta} \log L_T(\mathcal{K}, \theta) \quad \text{s.t.} \quad R(\theta) = 0, \quad (7)$$

where  $\hat{\mathcal{K}} := (\hat{\mathcal{K}}_1, \dots, \hat{\mathcal{K}}_G)$  with  $\hat{\mathcal{K}}_g := (\hat{k}_{g1}, \dots, \hat{k}_{gm})$  and  $\hat{\theta} := (\hat{\beta}, \hat{\Sigma})$  with  $\hat{\beta} := (\hat{\beta}_1, \dots, \hat{\beta}_{m+1})$  and  $\hat{\Sigma} := (\hat{\Sigma}_1, \dots, \hat{\Sigma}_{m+1})$ . Using the estimates of basic parameters  $\hat{\theta}$ , we can define  $\hat{\beta}_{gj}$  as in (4) and  $\hat{\theta}_{t,\mathcal{K}} := (\hat{\beta}_{t,\mathcal{K}}, \hat{\Sigma}_{t,\mathcal{K}})$  as in (6) given a collection of break dates  $\mathcal{K}$ .

The null hypothesis of common breaks in the model (2) can be stated as

$$H_0 : \mathcal{K}_g^0 = (T_1^0, \dots, T_m^0) \quad \text{for all } g = 1, \dots, G, \quad (8)$$

and the alternative hypothesis is simply the negation of the null hypothesis.<sup>7</sup> In what follows, we use  $\mathcal{T}$  to denote a collection of common break dates. More precisely, the collection  $\mathcal{T}$  is an element of the following set:

$$\Xi_{\nu, H_0} := \{ \mathcal{K} \in \Xi_\nu : \mathcal{K}_g = (T_1, \dots, T_m) \quad \text{for all } g = 1, \dots, G \}.$$

This set restricts break dates  $\mathcal{K}$  so that  $\mathcal{K}_1 = \dots = \mathcal{K}_G = (T_1, \dots, T_m)$  and  $T_j - T_{j-1} \geq \nu T$  for all  $j = 1, \dots, m$ . We also let  $\mathcal{T}^0$  be the collection of true break dates under common

---

<sup>7</sup> In Section 3.3, we discuss more details of the alternative hypothesis and consider the asymptotic power of the proposed test.



breaks assumption. The quasi-maximum likelihood estimates under the null can be obtained from the maximization problem with a restricted set of candidate break dates:

$$(\tilde{\mathcal{T}}, \tilde{\theta}) := \arg \max_{(\mathcal{T}, \theta) \in \Xi_{\nu, H_0} \times \Theta} \log L_T(\mathcal{K}, \theta) \quad \text{s.t.} \quad R(\theta) = 0, \quad (9)$$

where the break dates  $\tilde{\mathcal{T}}$  under the null hypothesis (8) consists of break dates  $\tilde{T}_1, \dots, \tilde{T}_m$ , all of which are separated by a positive fraction of the sample size. Without common breaks restrictions, however, the break date estimates  $\hat{\mathcal{K}}$  are simply allowed to be distinct but not necessarily separated by a positive fraction of the sample size across groups. This will be important since the setup of Qu and Perron (2007) require the maximization to be taken over asymptotically distinct elements and their proof for the convergence rate of the estimates relies on this premise. Hence, we will need to provide a detailed proof of the convergence rate under this less restrictive maximization problem (see Section 3).

The test considered is simply the likelihood ratio test that compares the values of the likelihood function with common breaks restriction and the one without the restriction. We define the test statistics as

$$CB_T := \log L_T(\hat{\mathcal{K}}, \hat{\theta}) - \log L_T(\tilde{\mathcal{T}}, \tilde{\theta}).$$

For asymptotic analysis of this test statistics, it is useful to employ a sort of normalization by using the log-likelihood function evaluated at the true parameters  $(\mathcal{K}^0, \theta^0)$  and we express

$$CB_T = \ell_T(\hat{\mathcal{K}}, \hat{\theta}) - \ell_T(\tilde{\mathcal{T}}, \tilde{\theta}),$$

where  $\ell_T(\mathcal{K}, \theta) := \log L_T(\mathcal{K}, \theta) - \log L_T(\mathcal{K}^0, \theta^0)$  for any  $(\mathcal{K}, \theta) \in \Xi_\eta \times \Theta$ . Since the two log likelihoods on which the test depends are maximized separately, we can consider asymptotic properties of each normalized log likelihood to obtain the limit distribution of the test.

### 2.3 Examples

Given that the notation is rather complex it is useful to illustrate the framework explained in the preceding subsection via examples.

**Example 1 (changes in intercepts):** We consider a two-equations system of autoregressive (AR) model with structural changes in intercepts:

$$\begin{aligned} y_{1t} &= \mu_{1j} + \alpha_1 y_{1,t-1} + u_{1t}, \\ y_{2t} &= \mu_{2j} + \alpha_2 y_{2,t-1} + u_{2t}, \end{aligned}$$

for  $T_{j-1} + 1 \leq t \leq T_j$  with  $j = 1, 2$ , where  $(u_{1t}, u_{2t})'$  have a covariance matrix  $\Sigma$ . In the above model, the basic parameters other than intercepts are assumed to be constant and a break in intercepts share the break date  $T_1$ . In the model (2), we set  $x_{tT} = (1, y_{1,t-1}, y_{2,t-1})'$ ,  $\beta_j = (\mu_{1j}, \alpha_{1j}, \mu_{2j}, \alpha_{2j})'$  and  $E[u_t u_t'] = \Sigma_j$ . We choose a selection matrix  $S = \langle s_{ij} \rangle$  as a  $6 \times 4$  matrix taking 1 at the entries  $s_{11}, s_{22}, s_{34}$  and  $s_{64}$  and 0 at the rest. Also we impose restrictions on basic parameters through the function  $R(\theta) = (\alpha_{11} - \alpha_{12}, \alpha_{21} - \alpha_{22}, \text{vec}(\Sigma_1) - \text{vec}(\Sigma_2))'$  in (3). On the other hand, the model (5) allow us to separate  $\beta_j$  into  $\beta_{1j} = (\mu_{1j}, \alpha_{1j}, 0, 0)'$  and  $\beta_{2j} = (0, 0, \mu_{2j}, \alpha_{2j})'$ , so that we can set  $\mathcal{G}_1 = \{1, 2\}$  and  $\mathcal{G}_2 = \{3, 4\}$ . We have two possibly distinct break dates  $k_{11}$  and  $k_{21}$  for parameter groups  $\{\beta_{1j}\}_{j=1}^2$  and  $\{(\beta_{2j}, \Sigma_j)\}_{j=1}^2$ , respectively. We address the issue of testing the null of  $H_0 : (k_{11}, k_{21}) = (T_1, T_1)$ .

**Example 2 (a single equation model):** Consider a single equation model:

$$y_{1t} = \mu + \alpha_j z_{1,t} + \gamma_j(t/T) + \delta_j T^{-1/2} w_{1t} + u_{1t}, \quad u_{1t} \sim N(0, \sigma_j^2),$$

for  $T_{j-1} + 1 \leq t \leq T_j$  with  $j = 1, 2, 3$ . In this example, the basic parameters other than intercepts have two structural changes. Under the model (2) with a break dates  $T_1$  and  $T_2$ , we set  $x_{tT} = (1, z_{1t}, t/T, T^{-1/2} w_{1t})'$ ,  $S = I_4$ ,  $\beta_j = (\mu_j, \alpha_j, \gamma_j, \delta_j)'$  and  $E[u_{1t}^2] = \sigma_j^2$ . Restrictions in (3) are imposed by the function  $R(\theta) = (\mu_1 - \mu_2, \mu_2 - \mu_3)'$ . We consider a test of common breaks against the alternative that all coefficients break dates are possibly not common, while a coefficient  $\delta_j$  and a variance  $\sigma_j^2$  share two break dates. In this case, we separate  $\beta_j$  into three vectors  $\beta_{1j} = (\mu_j, \alpha_j, 0, 0)$ ,  $\beta_{2j} = (0, 0, \gamma_j, 0)$  and  $\beta_{3j} = (0, 0, 0, \delta_j)$ . For these parameters groups, we assign a set of break dates  $\mathcal{K}_g = (k_{g1}, k_{g2})$  for  $g = 1, \dots, 3$  and we set  $\mathcal{G}_1 = \{1, 2\}$ ,  $\mathcal{G}_2 = \{3\}$  and  $\mathcal{G}_3 = \{4\}$ . The break dates for the last group,  $\mathcal{K}_3$ , are also the ones for the variance. This example shows that our framework can accommodate common breaks not only across equations in a system but also within an equation.

### 3 Asymptotic results

This section presents asymptotic results. We first provide convergence rates of estimates of break dates and basic parameters, allowing for the possibility that break dates of different basic parameters may not be asymptotically distinct. This condition is substantially less restrictive than the ones usually assumed in the existing literature and particularly includes the assumption of common breaks as a special case. Next, we provide the limiting distribution of the quasi-likelihood ratio test for common breaks under the null hypothesis. Finally we provide asymptotic power analysis of the test under a fixed alternative as well as a local alternative. Our result shows that the test has non-trivial asymptotic power.

### 3.1 The rate of convergence of the estimates.

We consider the case where we obtain the quasi-likelihood estimates  $(\hat{\mathcal{K}}, \hat{\theta})$  as in (7), using the observations  $\{(y_t, x_{tT})\}_{t=1}^T$  generated by the model in (5) with collections of true parameter values  $(\mathcal{K}^0, \theta^0)$ . The results presented in this subsection can apply for the estimates obtain under the null hypothesis since the model under the null is a special case of the current setup. To obtain asymptotic results, the following assumptions are employed.

#### Assumptions:

- A1.** There exists a constant  $k_0 > 0$  such that for all  $k > k_0$ , the minimum eigenvalues of matrices  $k^{-1} \sum_{t=s}^{s+k} x_{tT} x'_{tT}$  are bounded away from zero for every  $s = 1, \dots, T - k$ .
- A2.** Define a sigma-algebra  $\mathcal{F}_t := \sigma(\{z_s, u_{ws}, \eta_s\}_{s \leq t})$  for  $t \in \mathbb{Z}$ . (a) Define  $\zeta_t := (z'_t, u'_{wt})'$  and let  $z_t$  include a constant term. A sequence  $\{\zeta_t \otimes \eta_t, \mathcal{F}_t\}_{t \in \mathbb{Z}}$  forms a strongly mixing ( $\alpha$ -mixing) sequence with size  $-(4+\delta)/\delta$  for some  $\delta \in (0, 1/2)$  and satisfies that  $E[z_t \otimes \eta_t] = 0$  and  $\sup_{t \in \mathbb{Z}} \|\zeta_t \otimes \eta_t\|_{4+\delta} < \infty$ . (b) It is also assumed that  $\{\eta_t \eta'_t - I_n\}_{t \in \mathbb{Z}}$  satisfies the same mixing and moment conditions as in (a) of this assumption. (c) A sequence  $\{w_0 \otimes \eta_t\}_{t \in \mathbb{Z}}$  forms a strong mixing sequence as in (a) with  $\sup_{t \in \mathbb{Z}} \|w_0 \otimes \eta_t\|_{4+\delta} < \infty$  and the initial condition  $w_0$  is  $\mathcal{F}_0$ -measurable.
- A3.**  $k_{gj}^0 = \lceil T \lambda_{gj}^0 \rceil$  for every  $(g, j) \in \{1, \dots, G\} \times \{1, \dots, m\}$ , where  $0 < \lambda_{g1}^0 < \dots < \lambda_{gm}^0 < 1$ .
- A4.** For every parameter group  $g$  and regime  $j$ , there exists a  $p \times 1$  vector  $\delta_{gj}$  and an  $n \times n$  matrix  $\Phi_j$  such that  $\beta_{g,j+1}^0 - \beta_{gj}^0 = v_T \delta_{gj}$  and  $\Sigma_{j+1}^0 - \Sigma_j^0 = v_T \Phi_j$ , where both  $\delta_{gj}$  and  $\Phi_j$  are independent of  $T$ , and  $v_T > 0$  is a scalar satisfying  $v_T \rightarrow 0$  and  $\sqrt{T} v_T / \log T \rightarrow \infty$  as  $T \rightarrow \infty$ . Let  $\delta_j := \sum_{g=1}^G \delta_{gj}$  for  $j = 1, \dots, m + 1$ .
- A5.** The true basic parameters  $(\beta^0, \Sigma^0)$  belong to the compact parameter space

$$\Theta := \left\{ \theta : \max_{1 \leq j \leq m+1} \|\beta_j\| \leq c_1, \ c_2 \leq \min_{1 \leq j \leq m+1} \lambda_{\min}(\Sigma_j), \ \max_{1 \leq j \leq m+1} \lambda_{\max}(\Sigma_j) \leq c_3, \right\},$$

for some constants  $c_1 < \infty$ ,  $0 < c_2 \leq c_3 < \infty$ , where  $\lambda_{\min}(\cdot)$  and  $\lambda_{\max}(\cdot)$  denote the smallest and largest eigenvalues of a matrix in its argument, respectively.

Assumption A1 ensures that there is no local collinearity problem so that a standard invertibility requirement holds if the number of observations of some sub-sample is more than  $k_0$ , not depending on  $T$ . Assumption A2 determines the dependence structure of the processes  $\{\zeta_t \otimes \eta_t\}$ ,  $\{\eta_t \eta'_t - I_n\}$  and  $\{w_0 \otimes \eta_t\}$  to guarantee that they are short memory processes and have bounded fourth moments. The assumptions are imposed to obtain a

functional central limit theorem and a generalized Hájek and Rényi (1955) type inequality that allow us to derive their rate of convergence. Assumption A2 also specifies that the stationary regressors are contemporaneously uncorrelated with the errors and that a constant term is included in  $z_t$ . The former is a standard requirement to obtain consistent estimates and the latter is for notational simplicity since the results reported below are the same without a constant term.<sup>89</sup> It is important to note that no assumption is imposed on the correlation between the innovations to the  $I(1)$  regressors and the errors. Hence, we allow endogenous  $I(1)$  regressors. Assumption A3 implies asymptotically distinct breaks within each parameter group, but not necessarily across groups. Assumption A4 implies a shrinking shifts asymptotic framework whereby the magnitudes of the shifts converge to zero as the sample size increases. This condition is necessary to develop a limit distribution theory for break dates that does not depend on the exact distributions of the regressors and the errors, and this has been used in the literature (e.g., Bai, 1997; Bai and Perron, 1998; Bai et al., 1998).<sup>10</sup> Assumption A5 implies that the data are generated by a model with a finite conditional mean and innovations having the non-degenerate covariance matrix.

As stated above, the break dates are estimated from a set  $\Xi_\nu$ , which requires candidate break dates to be separated by some fraction of the sample size only within parameter groups. Thus, we cannot appeal to the results in Qu and Perron (2007) about the rate of convergence of the estimates, and a more general result is needed. The following theorem presents results about convergence rates of the estimates.

**Theorem 1.** *Suppose that Assumption A1-A5 hold. Then,*

(a) *for all  $(g, j) \in \{1, \dots, G\} \times \{1, \dots, m\}$ ,*

$$v_T^2(\hat{k}_{gj} - k_{gj}^0) = O_p(1),$$

(b) *for all  $(g, j) \in \{1, \dots, G\} \times \{1, \dots, m + 1\}$ ,*

$$\sqrt{T}(\hat{\beta}_{gj} - \beta_{gj}^0) = O_p(1) \quad \text{and} \quad \sqrt{T}(\hat{\Sigma}_j - \Sigma_j^0) = O_p(1).$$

This theorem establishes the usual rates of convergence obtained in Bai and Perron (1998), Bai et al. (1998) and Qu and Perron (2007) under our less restrictive set of assumptions with respect to the optimization problem as well as time-series properties of regressors.

---

<sup>8</sup> To allow for the correlation between the stationary regressors and the errors, one may use the two-stage least squares method if relevant instrumental variables are available (see Hall et al., 2012, for example).

<sup>9</sup> When a constant term is not included in  $z_t$  in contrast to Assumption A1, one additionally need to assume that the sequence  $\{\eta_t\}_{t \in \mathbb{Z}}$  satisfies the same mixing and moment conditions as in Assumption A1(a).

<sup>10</sup> Our condition that  $\sqrt{T}v_T/\log T \rightarrow 0$  as  $T \rightarrow \infty$  is the same as that in Bai et al. (1998) and weaker than the ones in Bai (1997) and Bai and Perron (1998), who requires  $\sqrt{T}v_T/T^\rho \rightarrow 0$  for some  $\rho \in (0, 1/2)$ .

The importance of these results is that they will allow us to analyze the properties of our test under compact sets for the parameters, namely, for some  $M > 0$ ,

$$\begin{aligned}\bar{\Xi}_M &:= \{\mathcal{K} \in \Xi_\nu : \max_{1 \leq g \leq G} \max_{1 \leq j \leq m} |k_{gj} - k_{gj}^0| \leq Mv_T^{-2}\} \\ \bar{\Theta}_M &:= \{\theta \in \Theta : \max_{1 \leq g \leq G} \max_{1 \leq j \leq m+1} \|\beta_{gj} - \beta_{gj}^0\| \leq MT^{-1/2}, \max_{1 \leq j \leq m+1} \|\Sigma_j - \Sigma_j^0\| \leq MT^{-1/2}\}.\end{aligned}$$

We also have a result that expresses the restricted likelihood in two parts: one that involves only the break dates and the true values of the coefficients; the other involving the true values of the break dates, the basic parameters and the restrictions. Thus, the estimates of the break dates are not affected by the restrictions imposed on the coefficients, while the limiting distributions of these estimates are influenced by the restrictions but not the estimation of the break dates.

**Theorem 2.** *Suppose that Assumption A1-A5 hold. Then,*

$$\sup_{(\mathcal{K}, \theta) \in \bar{\Xi}_M \times \bar{\Theta}_M} \ell_T(\mathcal{K}, \theta) = \sup_{\mathcal{K} \in \bar{\Xi}_M} \ell_T(\mathcal{K}, \theta^0) + \sup_{\theta \in \bar{\Theta}_M} \{\ell_T(\mathcal{K}^0, \theta) + \lambda' R(\theta)\} + o_p(1). \quad (10)$$

It implies that when analyzing asymptotic property of the break date estimates, one can ignore the restrictions in (3). This will prove especially convenient to obtain the limit distribution of our test. Since the quasi-likelihood ratio test can be expressed as a difference of two normalized log likelihoods evaluated at different break dates, the second term on the right-hand side of (10) is canceled out in the test statistics.

### 3.2 The limit distribution of the likelihood ratio test

We now establish the limit distribution of the quasi-likelihood ratio test under the null hypothesis of common breaks in (8). To this end, let the data consist of the observations  $\{(y_t, x_{tT})\}_{t=1}^T$  from the model of (2) with true basic parameters  $\theta^0 = (\beta^0, \Sigma^0)$  and true break dates  $\mathcal{T}^0$  consisting of  $T_1^0, \dots, T_m^0$ . Theorem 1(a) shows that, for all  $(g, j) \in \{1, \dots, G\} \times \{1, \dots, m\}$ , there exists a sufficiently large  $M$  such that  $|\hat{k}_{gj} - T_j^0| \leq Mv_T^{-2}$  and  $|\tilde{T}_j - T_j^0| \leq Mv_T^{-2}$  in probability, while the true break dates are asymptotically distinct. This implies that we can restrict our analysis to an interval centered at the true break  $T_j^0$  with interval length  $2Mv_T^{-2}$  for each regime  $j = 1, \dots, m$ . More precisely, given a sufficiently large  $M$ , we have that  $\theta_{t, \hat{\mathcal{K}}}^0 = \theta_{t, \mathcal{T}^0}^0$  and  $\theta_{t, \tilde{\mathcal{T}}}^0 = \theta_{t, \mathcal{T}^0}^0$  for all  $t \notin \cup_{j=1}^m [T_j^0 - Mv_T^{-2}, T_j^0 + Mv_T^{-2}]$ , in probability.

This together with Theorem 2 yields that, under the null of (8),

$$\begin{aligned}
CB_T &= \max_{\mathcal{K} \in \Xi_M} \sum_{j=1}^m \sum_{\underline{k}_j+1}^{\bar{k}_j} \left\{ \log f(y_t | X_{tT}, \theta_{t,\mathcal{K}}^0) - \log f(y_t | X_{tT}, \theta_{t,T^0}^0) \right\} \\
&\quad - \max_{\mathcal{T} \in \Xi_M} \sum_{j=1}^m \sum_{\underline{T}_j+1}^{\bar{T}_j} \left\{ \log f(y_t | X_{tT}, \theta_{t,\mathcal{T}}^0) - \log f(y_t | X_{tT}, \theta_{t,T^0}^0) \right\} + o_p(1),
\end{aligned}$$

where  $\bar{k}_j := \max\{k_{1j}, \dots, k_{Gj}, T_j^0\}$ ,  $\underline{k}_j := \min\{k_{1j}, \dots, k_{Gj}, T_j^0\}$ ,  $\bar{T}_j := \max\{T_j, T_j^0\}$  and  $\underline{T}_j := \min\{T_j, T_j^0\}$ . Since the neighborhoods of the true breaks expands as  $v_T$  decreases, an application of a FCLT for each neighborhood yields the limit distribution of the est. To this end, we make the following assumptions.

### Assumptions:

- B1.** Assumption A1 holds. As  $\Delta T_j^0 := (T_j^0 - T_{j-1}^0) \rightarrow \infty$ , the matrix  $(\Delta T_j^0)^{-1} \sum_{t=T_{j-1}^0+1}^{T_j^0} x_{tT} x_{tT}'$  converges to a random matrix not necessarily the same for all  $j = 1, \dots, m+1$ . Also,  $(\Delta T_j^0)^{-1} \sum_{t=T_{j-1}^0+1}^{T_j^0+[\Delta T_j^0]} z_t \xrightarrow{p} s\mu_{z,j}$  and  $(\Delta T_j^0)^{-1} \sum_{t=T_{j-1}^0+1}^{T_j^0+[\Delta T_j^0]} z_t z_t' \xrightarrow{p} sQ_{zz,j}$  uniformly in  $s \in [0, 1]$  as  $\Delta T_j^0 \rightarrow \infty$ , where  $Q_{zz,j}$  is a non-random positive definite matrix.
- B2.** Assumption A2 holds. Define  $S_{k,j}(\ell) := \sum_{t=T_{j-1}^0+1}^{T_j^0+[\Delta T_j^0]} (\zeta_t \otimes \eta_t)$  for  $k, \ell \in \mathbb{N}$  and for  $j = 1, \dots, m+1$ . (i) If  $\{\zeta_t \otimes \eta_t\}_{t \in \mathbb{Z}}$  is weakly stationary within each segment, then, for any vector  $e \in \mathbb{R}^{(q_z+q_w)n}$  with  $\|e\| = 1$ ,  $\text{var}(e' S_{k,j}(0)) \geq v(k)$  for some function  $v(k) \rightarrow \infty$  as  $k \rightarrow \infty$ . (ii) If  $\{\zeta_t \otimes \eta_t\}_{t \in \mathbb{Z}}$  is not weakly stationary within each segment, we additionally assume that there is a positive definite matrix  $\Omega = [\omega_{i,s}]$  such that for any  $i, s \in \{1, \dots, p\}$ , we have, uniformly in  $\ell$ ,  $|k^{-1} E[(S_{k,j}(\ell))_i (S_{k,j}(\ell))_s] - \omega_{i,s}| \leq k^{-\psi}$ , for some  $C > 0$  and for some  $\psi > 0$ . We also assume the same condition for  $\{\eta_t \eta_t' - I_n\}_{t \in \mathbb{Z}}$ .
- B3.**  $T_j^0 = \lceil T \lambda_j^0 \rceil$  for every  $j = 1, \dots, m$ , where  $0 < \lambda_1^0 < \dots < \lambda_m^0 < 1$ .
- B4.** For every regime  $j$ , there exists a  $p \times 1$  vector  $\delta_j$  and a  $n \times n$  matrix  $\Phi_j$  such that  $\beta_{j+1}^0 - \beta_j^0 = v_T \delta_j$  and  $\Sigma_{j+1}^0 - \Sigma_j^0 = v_T \Phi_j$ , where both  $\delta_j$  and  $\Phi_j$  are independent of  $T$ , and  $v_T > 0$  is a scalar satisfying  $v_T \rightarrow 0$  and  $\sqrt{T} v_T / \log T \rightarrow \infty$  as  $T \rightarrow \infty$ .
- B5.** Assumption A5 holds.
- B6.** Let  $V_{T,w}(r) := T^{-1/2} \sum_{t=1}^{\lceil Tr \rceil} u_{wt}$  for  $r \in [0, 1]$ .  $V_{T,w}(\cdot) \Rightarrow \mathbb{V}_w(\cdot)$ , where  $\mathbb{V}_w(\cdot)$  is a Wiener processes having a covariance function  $\text{cov}(\mathbb{V}_w(r), \mathbb{V}_w(s)) = (r \wedge s) \Omega_w$  for  $r, s \in [0, 1]$  with a positive definite matrix  $\Omega_w := \lim_{T \rightarrow \infty} \text{var}(T^{-1/2} \sum_{t=1}^T u_{wt})$ .

**B7.** For all  $1 \leq s, t \leq T$ , (a)  $E[(z_t \otimes \eta_t)w'_s] = 0$ , (b)  $E[(z_t \otimes \eta_t)\text{vec}(\eta_s \eta'_s)'] = 0$ , and (c)  $E[(u_{zt} \otimes \eta_t)\text{vec}(\eta_s \eta'_s)'] = 0$ .

Assumption B1 rules out trending variables in the stationary regressors  $z_t$ . Assumption B2 is mild in the sense that they allow for substantial conditional heteroskedasticity and autocorrelation. It can be shown to apply to a large class of linear processes including those generated by all stationary and invertible ARMA models. This assumption is useful to describe the asymptotic behavior of the test and in particular to characterize the limit distribution. Here, we introduce some processes used later. For each  $j = 1, \dots, m$ , let  $\mathbb{V}_{z\eta,j}^{(1)}(\cdot)$  and  $\mathbb{V}_{z\eta,j}^{(2)}(\cdot)$  be the Brownian motions defined on the space  $D[0, \infty)^{nq}$  with zero means and covariance functions given by, for  $l = 1, 2$  and for  $s_1, s_2 > 0$ ,

$$E[\mathbb{V}_{z\eta,j}^{(l)}(s_1)\mathbb{V}_{z\eta,j}^{(l)}(s_2)'] = (s_1 \wedge s_2) \lim_{T \rightarrow \infty} \text{var}(\bar{V}_{T,z\eta,j}^{(l)}),$$

where  $\bar{V}_{T,z\eta,j}^{(1)} := (\Delta T_j^0)^{-1/2} \sum_{t=T_{j-1}^0+1}^{T_j^0} (z_t \otimes \eta_t)$  and  $\bar{V}_{T,z\eta,j}^{(2)} := (\Delta T_{j+1}^0)^{-1/2} \sum_{t=T_j^0+1}^{T_{j+1}^0} (z_t \otimes \eta_t)$ . Similarly, define  $\mathbb{V}_{\eta\eta,j}^{(1)}(\cdot)$  and  $\mathbb{V}_{\eta\eta,j}^{(2)}(\cdot)$  as the Brownian motions defined on the space  $D[0, \infty)^{n^2}$  with zero means and covariance functions given by, for  $l = 1, 2$  and for  $s_1, s_2 > 0$ ,

$$E[\text{vec}(\mathbb{V}_{\eta\eta,j}^{(l)}(s_1))\text{vec}(\mathbb{V}_{\eta\eta,j}^{(l)}(s_2))'] = (s_1 \wedge s_2) \lim_{T \rightarrow \infty} \text{var}\{\text{vec}(\bar{V}_{T,\eta\eta,j}^{(l)})\},$$

where  $\bar{V}_{T,\eta\eta,j}^{(1)} := (\Delta T_j^0)^{-1/2} \sum_{t=T_{j-1}^0+1}^{T_j^0} (\eta_t \eta'_t - I_n)$  and  $\bar{V}_{T,\eta\eta,j}^{(2)} := (\Delta T_{j+1}^0)^{-1/2} \sum_{t=T_j^0+1}^{T_{j+1}^0} (\eta_t \eta'_t - I_n)$ . We define the two-sided Brownian motions as

$$\mathbb{V}_{z\eta,j}(s) := \begin{cases} \mathbb{V}_{z\eta,j}^{(1)}(-s), & s < 0 \\ 0, & s = 0 \\ \mathbb{V}_{z\eta,j}^{(2)}(s), & s > 0 \end{cases} \quad \text{and} \quad \mathbb{V}_{\eta\eta,j}(s) := \begin{cases} \mathbb{V}_{\eta\eta,j}^{(1)}(-s), & s < 0 \\ 0, & s = 0 \\ \mathbb{V}_{\eta\eta,j}^{(2)}(s), & s > 0. \end{cases}$$

Under Assumption B1,  $z_t$  is assumed to include a constant term and the process  $\mathbb{V}_{z\eta,j}^{(l)}(\cdot)$  includes some process depending purely on  $\{\eta_t\}$ . We denote it by  $\mathbb{V}_{\eta,j}^{(l)}(\cdot)$  for each  $l = 1, 2$  and also define a two-sided Brownian motion, denoted by  $\mathbb{V}_{\eta,j}(\cdot)$ , as before.

Assumption B3-B5 are similar to Assumption A3-A5 with some adjustments to satisfy restrictions under the null hypothesis. Assumption B6 requires the integrated regressors to follow a homogeneous distribution throughout the sample. Allowing for heterogeneity in the distribution of the errors underlying the  $I(1)$  regressors would be considerably more difficult, since we would, instead of having a limit distribution in terms of standard Wiener processes, have time-deformed Wiener processes according to the variance profile of the errors through time; see, e.g., Cavaliere and Taylor (2007). This would lead to important

complications given that, as shown below, the limit distribution of the estimates of the break dates depends on the whole time profile of the limit Wiener processes. It is possible to allow for trends in the  $I(1)$  regressors. The limiting distributions of the test to be derived will remain valid under different Wiener processes (see Hansen, 1992). The positive definiteness of the matrix  $\Omega_w$  rules out cointegration among the  $I(1)$  regressors and is needed to ensure a set of regressors that has a positive definite limit.

Assumption B7 restricts somewhat the class of models applicable but is quite mild by imposing sufficient, though not necessary, conditions that guarantee the independence of most Wiener processes described above, so that a limit distribution of the test is manageable. The condition (a) ensures that the autocovariance structure of  $I(0)$  regressors and the errors are uncorrelated with the  $I(1)$  variables. This is needed to guarantee that  $\mathbb{V}_{z\eta,j}(\cdot)$  are uncorrelated with  $\mathbb{V}_{w,j}(\cdot)$  and, being Gaussian, are therefore independent. Without these conditions, the analysis would be much more complex. Similarly, the condition (b) and (c) imply the independence between  $\mathbb{V}_{z\eta,j}(\cdot)$  and  $\mathbb{V}_{\eta\eta}(\cdot)$ .

In order to characterize the limit distribution of  $CB_T$  it is useful to first state some preliminary results about the limit distribution of some quantities. For  $s \in \mathbb{R}$  and for  $j = 1, \dots, m$ , let  $\bar{T}_j(s) := \max\{T_j(s), T_j^0\}$  and  $\underline{T}_j(s) := \min\{T_j(s), T_j^0\}$  where  $T_j(s) := T_j^0 + [sv_T^{-2}]$ . We define, for  $s, s_{Gj} \in \mathbb{R}$ ,

$$B_{T,j}(s, s_G) := v_T^2 \sum_{t=\underline{T}_j^0(s)+1}^{\bar{T}_j(s)} X_{tT}(\Sigma_{t,\mathcal{K}}^0)^{-1} X'_{tT} \quad \text{and} \quad W_{T,j}(s, s_{Gj}) := v_T \sum_{t=\underline{T}_j^0(s)+1}^{\bar{T}_j(s)} X_{tT}(\Sigma_{t,\mathcal{K}}^0)^{-1} u_t.$$

**Lemma 1.** *Suppose that Assumption B1-B7 hold. Then,*

$$\{B_{T,j}(\cdot, \cdot), W_{T,j}(\cdot, \cdot)\}_{j=1}^m \Rightarrow \{\mathbb{B}_j(\cdot, \cdot), \mathbb{W}_j(\cdot, \cdot)\}_{j=1}^m,$$

where

$$\begin{aligned} \mathbb{B}_j(s, s_G) &:= |s| S' \mathbb{D}_j(s) \otimes (\Sigma_{j+1}^0 \mathbb{1}_{\{s_G \leq s\}})^{-1} S \\ &\quad - \mathbb{1}_{\{|s_G| \leq |s|\}} |s_G| S' \mathbb{D}_j(s) \otimes \{(\Sigma_{j+1}^0)^{-1} - (\Sigma_j^0)^{-1}\} S, \end{aligned}$$

and

$$\begin{aligned} \mathbb{W}_j(s, s_G) &:= S' (I_q \otimes (\Sigma_{j+1}^0 \mathbb{1}_{\{s_G \leq s\}})^{-1}) \mathbb{V}_j(s) \\ &\quad - \text{sgn}(s_G) \mathbb{1}_{\{|s_G| \leq |s|\}} S' [I_q \otimes \{(\Sigma_{j+1}^0)^{-1} - (\Sigma_j^0)^{-1}\}] \mathbb{V}_j(s_G), \end{aligned}$$



with  $\mathbb{V}_j(s) := [\mathbb{V}_{z\eta,j}(s)', \varphi(\lambda_j^0)' \otimes \mathbb{V}_{\eta,j}(s)', \mathbb{V}_w(\lambda_j^0)' \otimes \mathbb{V}_{\eta,j}(s)']'$  and

$$\mathbb{D}_j(s) := \begin{pmatrix} Q_{zz,j+1\{0<s\}} & \mu_{z,j+1\{0<s\}} \varphi(\lambda_j^0)' & \mu_{z,j+1\{0<s\}} \mathbb{V}_w(\lambda_j^0)' \\ \varphi(\lambda_j^0) \mu'_{z,j+1\{0<s\}} & \varphi(\lambda_j^0) \varphi(\lambda_j^0)' & \varphi(\lambda_j^0) \mathbb{V}_w(\lambda_j^0)' \\ \mathbb{V}_w(\lambda_j^0) \mu'_{z,j+1\{0<s\}} & \mathbb{V}_w(\lambda_j^0) \varphi(\lambda_j^0)' & \mathbb{V}_w(\lambda_j^0) \mathbb{V}_w(\lambda_j^0)' \end{pmatrix}.$$

The theorem below presents the main result of the paper concerning the limit distribution of the test statistic. The result shows that the limit distribution can be expressed as the difference of maximums of the limit process without and with restrictions implied by the assumption of common breaks.

**Theorem 3.** *Let  $\mathbf{s}_j = (s_{1j}, \dots, s_{Gj})$  for  $j = 1, \dots, m$  and let  $\mathbf{1}$  is a  $G \times 1$  vector having 1 at all entries. Suppose Assumption B1-B7 hold. Then, under the null hypothesis of (8),*

$$CB_T \xrightarrow{d} CB_\infty := \sup_{\mathbf{s}_1, \dots, \mathbf{s}_m} \sum_{j=1}^m CB_\infty^{(j)}(\mathbf{s}_j) - \sup_{\mathbf{s}_1, \dots, \mathbf{s}_m} \sum_{j=1}^m CB_\infty^{(j)}(s_j \cdot \mathbf{1}),$$

where

$$\begin{aligned} CB_\infty^{(j)}(\mathbf{s}_j) &:= \frac{1}{2} \text{tr}(\Pi_j(s_{Gj}) \mathbb{V}_{\eta\eta,j}(s_{Gj})) + \frac{|s_{Gj}|}{4} \text{tr}(\{\Pi_j(s_{Gj})\}^2) - \sum_{g=1}^G \text{sgn}(s_{gj}) \delta'_{gj} \mathbb{W}_j(s_{gj}, s_{Gj}) \\ &\quad - \frac{1}{2} \sum_{g=1}^G \sum_{l=1}^G \delta'_{gj} \left\{ \mathbb{1}_{\{s_{gj} \vee s_{lg} \leq 0\}} \mathbb{B}_j(s_{gj} \vee s_{lg}, s_{Gj}) + \mathbb{1}_{\{0 < s_{gj} \wedge s_{lg}\}} \mathbb{B}_j(s_{gj} \wedge s_{lg}, s_{Gj}) \right\} \delta_{lj}, \end{aligned}$$

with

$$\Pi_j(s_{Gj}) := \begin{cases} (\Sigma_j^0)^{-1/2} \Phi_j(\Sigma_{j+1}^0)^{-1} (\Sigma_j^0)^{1/2}, & \text{if } s_{Gj} \leq 0 \\ -(\Sigma_{j+1}^0)^{-1/2} \Phi_j(\Sigma_j^0)^{-1} (\Sigma_{j+1}^0)^{1/2}, & \text{if } s_{Gj} > 0. \end{cases}$$

The limit distribution is quite complex and depends on nuisance parameters. However, they can be consistently estimated and it is easy to show that the coverage rates will be asymptotically valid provided  $\sqrt{T}$ -consistent estimates are used instead of the true values. The various quantities can be estimated as follows: for  $\Delta \hat{T}_j := \hat{T}_j - \hat{T}_{j-1}$ ,  $\hat{Q}_{zz,j} = (\Delta \hat{T}_j)^{-1} \sum_{t=\hat{T}_{j-1}+1}^{\hat{T}_j} z_t z_t'$ ,  $\hat{\mu}_{z,j} = (\Delta \hat{T}_j)^{-1} \sum_{t=\hat{T}_{j-1}+1}^{\hat{T}_j} z_t$ ,  $\hat{\delta}_{i,j} = \sum_{l \in \mathcal{G}_g} e_l \circ (\hat{\beta}_{j+1} - \hat{\beta}_j)$ ,  $\hat{\Sigma}_j = (\Delta \hat{T}_j)^{-1} \sum_{t=\hat{T}_{j-1}+1}^{\hat{T}_j} \hat{u}_t \hat{u}_t'$  and  $\hat{\Phi}_j = \hat{\Sigma}_{j+1} - \hat{\Sigma}_j$ . Also, the estimate of the long run variance of  $\{z_t \otimes \eta_t\}$  and  $\{\eta_t \eta_t' - I_n\}$  can be constructed using a method based on a weighted sum of sample autocovariances of the relevant quantities, as discussed in Andrews (1991) for instance. Though only  $\sqrt{T}$ -consistent estimates of  $(\beta, \Sigma)$  are needed, it is likely that more precise

estimates of these parameters will lead to better finite sample coverage rates. Hence, it is recommended to use the estimates obtained imposing the restrictions in (3) even though imposing restrictions does not have a first-order effect on the limiting distribution of the estimates of the break dates. As an immediate corollary, one can use the limit distribution to test the null hypothesis of common breaks for a specific regime, rather than all regimes. Similarly, the above result can be easily extend to test the hypothesis that common break dates are imposed on a part of parameter groups, while the break dates of the other groups are not necessarily common. We illustrate the test for common breaks in (8) and its variant through an application in Section 5.

As discussed in Section 1, there is one additional layer of difficulty compared to Bai and Perron (1998) or Qu and Perron (2007). In their analysis, the limit distribution can be evaluated using a closed form solution after some transformation, while no such solution is available here and thus we obtain critical values through simulation. This involves first simulating the Wiener processes appearing in the various Brownian motion processes by partial sums of *i.i.d.* normal random vectors (independent of each others given Assumption B7). One can then evaluate one realization of the limit distribution by replacing unknown values by their estimates as stated above. The procedure is then repeated many times to obtain the relevant quantiles. While conceptually straightforward, this procedure is nevertheless very computationally intensive. The reason is that for each replication we need to search over many possible combinations of all the permutations of the locations of the break dates. The procedure suggested is nevertheless quick enough to be feasible for common applications involving testing for few common break dates but the computational burden will have an impact on the extent to which we will be able to evaluate the adequacy of the limit distribution in finite samples via Monte Carlo simulations. Also, when testing for many common break dates, the computational burden increases exponentially with the number of common breaks being tested. Methods to alleviate this issue is an important avenue for further research.

### 3.3 Asymptotic power analysis

In this subsection, we provide an asymptotic power analysis of the test statistic  $CB_T$  when we use a critical value  $c_\alpha^*$  at the significance level  $\alpha$  from the asymptotic null distribution  $CB_\infty$ . As a fixed alternative hypothesis, we consider, for some  $\delta > 0$

$$H_1 : \max_{1 \leq g_1, g_2 \leq G} |k_{g_1, j}^0 - k_{g_2, j}^0| \geq \delta T \text{ for some } j = 1, \dots, m. \quad (11)$$

Given that  $k_{gj}^0 = [T\lambda_{gj}^0]$  for  $(g, j) \in \{1, \dots, G\} \times \{1, \dots, m\}$  under Assumption A3, the above condition is asymptotically equivalent to that  $\max_{1 \leq g_1, g_2 \leq G} |\lambda_{g_1 j}^0 - \lambda_{g_2 j}^0| \geq \delta$  for some  $j = 1, \dots, m$ , and thus can be considered as the fixed alternative hypothesis in term of break

fractions. As a local alternative hypothesis, we consider

$$H_{1T} : \max_{1 \leq g_1, g_2 \leq G} |k_{g_1, j}^0 - k_{g_2, j}^0| \geq M v_T^{-2} \text{ for some } j = 1, \dots, m, \quad (12)$$

for some constant  $M > 0$ , where  $v_T$  satisfies the condition in Assumption A4. We can also express (12) as that  $\max_{1 \leq g_1, g_2 \leq G} |\lambda_{g_1, j}^0 - \lambda_{g_2, j}^0| \geq M(\sqrt{T}v_T)^{-2}$  for some  $j = 1, \dots, m$ . The following theorem shows that the proposed test statistic is consistent against the fixed alternative and also has non-trivial local power against the local alternatives.

**Theorem 4.** *Let  $c_\alpha^* := \inf \{c \in \mathbb{R} : \Pr\{CB_\infty \leq c\} \geq 1 - \alpha\}$ . Suppose that Assumptions A1-A5 hold. Then, (a) under the fixed alternative in (11) with any  $\delta \in (0, 1]$ ,*

$$\lim_{T \rightarrow \infty} \Pr \{CB_T > c_\alpha^*\} = 1,$$

*and (b) under the local alternative in (12) with a sufficiently large  $M$ ,*

$$\lim_{T \rightarrow \infty} \Pr \{CB_T > c_\alpha^*\} = 1.$$

The result (a) in the above theorem shows that the test proposed here is consistent against the fixed alternative that break dates are asymptotically distinctive. Moreover, the result (b) states that the test has non-trivial asymptotic power even against the local alternative.

## 4 Monte Carlo simulations

In this section, we provide simulation results regarding the finite sample performance of the test proposed in terms of size and power. Given the very high computational cost of the simulations, we have to choose a simple setup and focus on limited cases (small sample size and a few parameter values).<sup>11</sup> Hence, the results should be interpreted in light of this. We here adopt a similar setup to the one used in Bai et al. (1998) in that we use an autoregressive system with changes in intercepts. The data generating process considered is the following bivariate system with a single break in each equation:

$$\begin{aligned} y_{1t} &= \mu_{1j} + \alpha y_{1,t-1} + u_{1t} \\ y_{2t} &= \mu_{2j} + \alpha y_{2,t-1} + u_{2t}, \end{aligned}$$

for  $T_{j-1} + 1 \leq t \leq T_j$  with  $j = 1, 2$ , where  $(u_{1t}, u_{2t})' \sim i.i.d. N(0, I_2)$ . Hence, only the intercept is allowed to change at some date  $T_1$  under the null hypothesis of common breaks

---

<sup>11</sup> For fixed parameter values and for each replication out of 500 realizations, it takes roughly two hours or more to simulate our critical values for the common break test. The estimation is also time consuming because we need to search over many possible combinations of all the permutations of two break locations.

and at some dates  $k_{1i}$  for the  $i^{th}$  equation under the alternative. The number of observations is set to  $T = 100$ . We use 500 replications and to generate the critical values we use 3,000 repetitions. Results are reported for the values of the autoregressive parameter  $\alpha = 0.0, 0.4$  and  $0.8$ . The break date is set at mid-sample, i.e.,  $T_1^0 = 50$ , and a trimming parameter  $\nu$  is set to be  $0.15$ .

We examine empirical rejection frequencies under the null. We set  $\mu_{i1} = 1$  and let  $\delta_i := \mu_{i2} - \mu_{i1}$  denote the magnitude of mean shift taking a value of  $\{0.50, 0.75, 1.00, 1.25, 1.50\}$  for each equation. The result is reported in Table 1, given nominal sizes of 10%, 5% and 1%. First, when the autoregressive process has no or moderate dependency ( $\alpha = 0.0$  or  $\alpha = 0.4$ ), the empirical size of the test is either conservative or close to the nominal size. Given the small sample size, this size property is satisfactory. When the autoregressive parameter is close to the boundary of the non-stationary region such that  $\alpha = 0.8$ , as expected, there are some liberal size distortions. When the magnitudes of the breaks are small, the test tends to over-reject the null hypothesis. This is due to the fact that for very small breaks the break date estimates are quite imprecise and are more likely to be affected by the highly dependent series than break sizes, so that the test depends on the log likelihoods evaluated at out of neighborhoods of true break dates. Once magnitude of break sizes increases enough, however, the size of the test quickly approaches the nominal level. These results are encouraging given that the sample size considered is small.

To analyze power, we also set  $\mu_{i1} = 1$ , while we consider a value of  $\{0.50, 1.00, 1.50\}$  as the magnitude of mean shift. The break date in the first equation is kept fixed at  $k_1 = 35$ , while the break date in the second equation takes values  $k_2 = 35, 40, 45, 50, 55$ . The power is a function of the difference between the break dates,  $k_2 - k_1$ . The results are presented in Figure 1, where the horizontal axis in each box represents the difference  $k_2 - k_1$  and the vertical axis shows the empirical rejection frequency. As before, when the magnitudes of the breaks are small, the estimates of the break dates are quite imprecise so that the data is not informative enough to reject common breaks assumption and the test has little power. However, when the magnitudes of the changes reach 1, the power increases rapidly as the distance between the break dates increases. The results are qualitatively similar for the values of  $\alpha$  considered.

## 5 Application

In this section, we apply the statistical test for common breaks to inflation series, following the study by Clark (2006). He analyzes the persistence of a number of disaggregated inflation series by employing the sum of the autoregressive (AR) coefficients in the AR model as a

persistence measure, and documents that the persistence is very high and close to one without allowing for a mean shift in the model, whereas the persistence declines substantially when allowing for a mean shift. Although the decline of the persistent in a mean shift model has been documented in the literature (e.g. Perron, 1990), he founds that this decline of the persistence is more pronounced among the disaggregated measures compared to the various aggregate measures. The issue of importance is that Clark (2006) assumes a common mean shift in the AR model, following Bai et al. (1998) but its validity has not yet been examined.

We consider a subset of the series analyzed in Clark (2006), namely the inflation measures for durables, nondurables and services. These are taken from the NIPA accounts and cover the period 1984-2002 at the quarterly frequency, see Clark (2006) for more details. Let  $\{(y_{1t}, y_{2t}, y_{3t})\}_{t=1}^T$  denote the inflation series of durables, nondurables and services and consider the AR model with a mean shift for each series  $i = 1, 2, 3$ :

$$y_{it} = \mu_i + \delta_i \mathbb{1}_{\{k_i+1 \leq t\}} + \alpha_i^{(1)} y_{i,t-1} + \cdots + \alpha_i^{(p_i)} y_{i,t-p_i} + u_{it}, \quad t = 1, \dots, T,$$

where  $\mu_i$  is an intercept parameter,  $\delta_i$  is a break size with  $k_i$  being a break date, the parameters,  $\alpha_i^{(1)}, \dots, \alpha_i^{(p_i)}$ , are AR coefficients with  $p_i$  denoting the lag length and  $u_{it}$  is an error term. The persistence of each series is measured by the sum  $\alpha_i^{(1)} + \cdots + \alpha_i^{(p_i)}$  for  $i = 1, 2, 3$ . Clark (2006) uses the Akaike information criterion (AIC) to select the AR lag length such that  $(p_1, p_2, p_3) = (4, 5, 3)$  and also presents some evidence to support a mean shift in the AR models by applying break tests for each series and multiple series.

We present our empirical results in Table 2. We first replicate a part of the results in Clark (2006). We find that when not allowing for a mean shift, the persistence measure is indeed quite high ranging from 0.855 to 0.921. It is also documented that the persistence measure decreases to a large extent for non-durables and services but not so much for durables when the common break date in 1993:Q1 for the intercept is imposed. The common break date 1993:Q1 is not estimated but treated as a known break date in Clark (2006). When we use the Seemingly Unrelated Regressions (SUR) method with an unknown common break, following Bai et al. (1998), the point estimates are similar expect that the break date is estimated at 1992:Q1.

We now use our test to assess the validity of common breaks assumption. In Table 2, we report values of the test statistic for several null hypotheses as well as critical values corresponding to a significance level 5%, each of which is obtained through simulation with 3,000 repetitions. First, we consider the null hypothesis of common breaks in three inflation series, i.e.,  $H_0 : k_1 = k_2 = k_3$ . The value of the test statistic obtained is 9.051 and the critical value is 3.858, so that the test rejects the null hypothesis of common breaks at the 5% significance level. Next, we test the null hypothesis of common breaks in two inflation

series under a system of three inflation series, separately. That is, we calculate the test statistics for each null,  $H_0 : k_1 = k_2$ ,  $H_0 : k_1 = k_3$ , and  $H_0 : k_2 = k_3$ . The values of the test statistics are 9.735 and 7.684 given corresponding critical values 1.871 and 3.003 for  $H_0 : k_1 = k_2$  and  $H_0 : k_1 = k_3$ , respectively, and thus each of these hypotheses is rejected at the 5% significance level. On the other hand, the value of the statistic for  $H_0 : k_2 = k_3$  is 0.749 given the critical value of 2.762. Thus, we cannot reject the null of common breaks in the nondurables and service series.

We then estimate a system of three inflation series imposing the common break only in the nondurables and service series (i.e.,  $k_2 = k_3$ ). For the nondurables and services series, the common break date is estimated at 1992:Q1, which is the same as when allowing for an unknown common break date in all series, and the parameter estimates are also broadly similar. Things are quite different for the durables series. In this case, the estimate of the break date is 1995:Q1. What is interesting is that the decrease in persistence is very important with an estimate of 0.324 compared to an estimate of 0.805 obtained assuming a common break date across the three series. Hence, allowing for different break dates for durables series and the others (nondurables and service), we document the substantial declines in the persistent measure across all three series. Moreover, we report the 95% confidence intervals for the estimated break dates: [1994:Q2, 1995:Q4] and [1991:Q3, 1992:Q3] for durables series and the others, respectively. These non-overlap intervals are consistent with our results.

## 6 Conclusion

This paper provides a procedure to test for common breaks across or within equations. Our framework is very general and allows integrated regressors and trends as well as stationary regressors. The test considered is the quasi-likelihood ratio test assuming normal errors, though as usual the limit distribution of the test remains valid with non-normal errors. Also of independent interest, we provide results about the rate of convergence when searching over all possible partitions subject only to the requirement that each regime of different parameters contains at least as many observations as some positive fraction of the sample size. Simulation results showed that the test has good finite sample properties. We also provide an application to various measures of inflation to illustrate its usefulness. As mentioned, the implementation of the test is very computationally intensive. When testing for many common breaks, the computational burden increases exponentially with the number of common breaks being tested so that the burden can be very large in some applications. Methods to alleviate this problem is an important avenue for further research that we intend to pursue.

## References

- ANDREWS, D. W. K. (1991): “Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation,” *Econometrica*, 59, 817–58.
- (1993): “Tests for Parameter Instability and Structural Change with Unknown Change Point,” *Econometrica*, 61, 821–56.
- BAI, J. (1994): “Least Squares Estimation of a Shift in Linear Processes,” *Journal of Time Series Analysis*, 15, 453–472.
- (1995): “Least Absolute Deviation Estimation of a Shift,” *Econometric Theory*, 11, 403–436.
- (1997): “Estimation of a Change Point in Multiple Regression Models,” *The Review of Economics and Statistics*, 79, 551–563.
- (1998): “Estimation of Multiple-Regime Regressions with Least Absolutes Deviation,” *Journal of Statistical Planning and Inference*, 74, 103–134.
- (2000): “Vector Autoregressive Models with Structural Changes in Regression Coefficients and in Variance-Covariance Matrices,” *Annals of Economics and Finance*, 1, 303–339.
- (2010): “Common Breaks in Means and Variances for Panel Data,” *Journal of Econometrics*, 157, 78–92.
- BAI, J., R. L. LUMSDAINE, AND J. H. STOCK (1998): “Testing For and Dating Common Breaks in Multivariate Time Series,” *Review of Economic Studies*, 65, 395–432.
- BAI, J. AND P. PERRON (1998): “Estimating and Testing Linear Models with Multiple Structural Changes,” *Econometrica*, 66, 47–78.
- (2003): “Computation and Analysis of Multiple Structural Change Models,” *Journal of Applied Econometrics*, 18, 1–22.
- BALTAGI, B. H., Q. FENG, AND C. KAO (2015): “Estimation of Heterogeneous Panels with Structural Breaks,” forthcoming, *Journal of Econometrics*, Nanyang Technological University.
- BREITUNG, J. AND S. EICKMEIER (2011): “Testing for Structural Breaks in Dynamic Factor Models,” *Journal of Econometrics*, 163, 71–84.
- CAVALIERE, G. AND A. R. TAYLOR (2007): “Testing for Unit Roots in Time Series Models with Non-stationary Volatility,” *Journal of Econometrics*, 140, 919–947.
- CHENG, X., Z. LIAO, AND F. SCHORFHEIDE (2016): “Shrinkage Estimation of High-Dimensional Factor Models with Structural Instabilities,” *The Review of Economic Studies*, rdw005.
- CLARK, T. E. (2006): “Disaggregate Evidence on the Persistence of Consumer Price Inflation,” *Journal of Applied Econometrics*, 21, 563–587.
- CORRADI, V. (1999): “Deciding between  $I(0)$  and  $I(1)$  via FLIL-Based Bounds,” *Econometric Theory*, 15, 643–663.

- CORRADI, V. AND N. R. SWANSON (2006): “Bootstrap Conditional Distribution Tests in the Presence of Dynamic Misspecification,” *Journal of Econometrics*, 133, 779–806.
- (2014): “Testing for Structural Stability of Factor Augmented Forecasting Models,” *Journal of Econometrics*, 182, 100–118.
- CSÖRGÖ, M. AND L. HORVÁTH (1997): *Limit Theorems in Change-Point Analysis*, vol. 18, John Wiley & Sons Inc.
- DAVIDSON, J. (1994): *Stochastic Limit Theory: An Introduction for Econometricians*, Oxford University Press, UK.
- EBERLEIN, E. (1986): “On Strong Invariance Principles under Dependence Assumptions,” *The Annals of Probability*, 14, 260–270.
- EO, Y. AND J. MORLEY (2015): “Likelihood-Ratio-Based Confidence Sets for the Timing of Structural Breaks,” *Quantitative Economics*, 6, 463–497.
- HÁJEK, J. AND A. RÉNYI (1955): “Generalization of an Inequality of Kolmogorov,” *Acta Mathematica Hungarica*, 6, 281–283.
- HALL, A. R., S. HAN, AND O. BOLDEA (2012): “Inference regarding Multiple Structural Changes in Linear Models with Endogenous Regressors,” *Journal of Econometrics*, 170, 281–302.
- HALL, A. R. AND A. SEN (1999): “Structural Stability Testing in Models Estimated by Generalized Method of Moments,” *Journal of Business & Economic Statistics*, 17, 335–48.
- HAN, X. AND A. INOUE (2015): “Tests for Parameter Instability in Dynamic Factor Models,” *Econometric Theory*, 31, 1117–1152.
- HANSEN, B. E. (1992): “Tests for Parameter Instability in Regressions with I(1) Processes,” *Journal of Business & Economic Statistics*, 10, 321–35.
- HATANAKA, M. AND H. YAMADA (2003): *Co-trending*, Springer.
- HENDRY, D. F. AND G. E. MIZON (1998): “Exogeneity, Causality, and Co-breaking in Economic Policy Analysis of a Small Econometric Model of Money in the UK,” *Empirical Economics*, 23, 267–294.
- IBRAGIMOV, I. A. (1962): “Some Limit Theorems for Stationary Processes,” *Theory of Probability & Its Applications*, 7, 349–382.
- KEJRIWAL, M. AND P. PERRON (2008): “The Limit Distribution of the Estimates in Cointegrated Regression Models with Multiple Structural Changes,” *Journal of Econometrics*, 146, 59–73.
- KIM, D. (2011): “Estimating a Common Deterministic Time Trend Break in Large Panels with Cross sectional Dependence,” *Journal of Econometrics*, 164, 310–330.
- KUROZUMI, E. AND P. TUVAANDORJ (2011): “Model Selection Criteria in Multivariate Models with Multiple Structural Changes,” *Journal of Econometrics*, 164, 218–238.
- OKA, T. AND Z. QU (2011): “Estimating Structural Changes in Regression Quantiles,” *Journal of Econometrics*, 162, 248–267.



- PERRON, P. (1990): “Testing for a Unit Root in a Time Series with a Changing Mean,” *Journal of Business & Economic Statistics*, 8, 153–62.
- PERRON, P. AND Y. YAMAMOTO (2014): “A Note on Estimating and Testing for Multiple Structural Changes in Models with Endogenous Regressors via 2SLS,” *Econometric Theory*, 30, 491–507.
- PERRON, P. ET AL. (2006): “Dealing with Structural Breaks,” *Palgrave handbook of econometrics*, 1, 278–352.
- QIAN, J. AND L. SU (2016): “Shrinkage Estimation of Common Breaks in Panel Data Models via Adaptive Group Fused Lasso,” *Journal of Econometrics*, 191, 86–109.
- QU, Z. (2008): “Testing for Structural Change in Regression Quantiles,” *Journal of Econometrics*, 146, 170–184.
- QU, Z. AND P. PERRON (2007): “Estimating and Testing Structural Changes in Multivariate Regressions,” *Econometrica*, 75, 459–502.
- SU, L. AND Z. XIAO (2008): “Testing for Parameter Stability in Quantile Regression Models,” *Statistics & Probability Letters*, 78, 2768–2775.
- YAMAMOTO, Y. AND S. TANAKA (2015): “Testing for Factor Loading Structural Change under Common Breaks,” *Journal of Econometrics*, 189, 187–206.

# Appendix

Through Appendix, we use  $C, C_1, C_2, \dots$  to denote generic positive constants without further clarification. Also, we use  $\text{diag}(\cdot)$  to denote the operator that sets all the off-diagonal elements of a matrix to zero and place its inputs on diagonal elements of a matrix.

The key ingredients in the proofs are a Strong Approximation Theorem (SAT), a Functional Central Limit Theorem (FCLT) and a generalized Hajek-Renyi inequality. We first state two technical lemmas.

**Lemma A.1.** *Let  $\{\varsigma_t\}_{t \in \mathbb{Z}}$  be a sequence of mean-zero,  $\mathbb{R}^d$ -valued random vectors satisfying Assumption B2. Define  $S_k(\ell) = \sum_{t=\ell+1}^{\ell+k} \varsigma_t$ , then, (a) (SAT) the covariances of  $k^{-1/2}S_k(\ell)$ ,  $\Omega_k$ , converge, with the limit denoted by  $\Omega$ , and there exists a Brownian Motion  $(W(t))_{t \geq 0}$  with covariance matrix  $\Omega$  such that  $\sum_{i=1}^t \varsigma_i - W(t) = O_{a.s.}(t^{1/2-\kappa})$  for some  $\kappa > 0$ ; (b) (FCLT)  $T^{-1/2} \sum_{t=1}^{[Tr]} \varsigma_t \Rightarrow \Omega^{1/2}W^*(r)$ , where  $W^*(r)$  is a  $\mathbb{R}^d$ -valued vector of independent Wiener processes and  $\Rightarrow$  denotes weak convergence under the Skorohod topology.*

The above lemme is proved in Lemma A.1 of Qu and Perron (2007), who use Theorem 2 in Eberlein (1986) together with the argument of Corradi (1999). The following lemma is an extension of the Hajek-Renyi inequality.

**Lemma A.2.** *Suppose that Assumption A1, A2 and A5 hold. Let  $\{b_k\}_{k \in \mathbb{N}}$  be a sequence of positive, non-increasing constants and let  $\{\xi_{tT}\}$  denote either  $\{X_{tT}\Sigma_{t,K}^{-1}u_t\}$  or  $\{\eta_t\eta'_t - I_n\}$ . Then, for any  $B > 0$  and for any  $k_1, k_2 \in \mathbb{N}$  with  $k_1 < k_2$ ,*

$$\Pr \left\{ \sup_{k_1 \leq k \leq k_2} \frac{1}{kb_k} \left\| \sum_{t=1}^k \xi_{tT} \right\| > B \right\} \leq \frac{C}{B^2} \left( \frac{1}{k_1 b_{k_1}^2} + \sum_{k=k_1+1}^{k_2} \frac{1}{(kb_k)^2} \right).$$

**Proof.** The assertion is proved if we show that  $\{X_{tT}\Sigma_{t,K}^{-1}u_t\}$  and  $\{\eta_t\eta'_t - I_n\}$  satisfy the  $L^2$ -mixingale condition in Lemma A6 of Bai and Perron (1998), which shows the Hajek-Renyi inequality for a  $L^2$ -mixingale sequence.<sup>12</sup> We shall consider only the case of  $\{X_{tT}\Sigma_{t,K}^{-1}u_t\}$  because the proof of the other case is simpler. We denote by  $E_t(\cdot) := E(\cdot | \mathcal{F}_t)$  for  $t \in \mathbb{Z}$ .

We can write  $X_{tT}\Sigma_{t,K}^{-1}u_t = S'(I_q \otimes \Sigma_{t,K}^{-1}(\Sigma_{t,K}^0)^{1/2})(x_{tT} \otimes \eta_t)$ , where  $\|S'(I_q \otimes \Sigma_{t,K}^{-1}(\Sigma_{t,K}^0)^{1/2})\| \leq C_1$  from Assumption A5 and the term  $(x_{tT} \otimes \eta_t)$  is  $\mathcal{F}_t$ -measurable. Thus, it suffices to show that there exist non-negative constants  $\{\psi_j\}_{j \geq 0}$  such that, for all  $t \geq 1$  and  $j \geq 0$ ,

$$\|E_{t-j}(x_{tT} \otimes \eta_t) - E(x_{tT} \otimes \eta_t)\|_2 \leq C_2 \psi_j, \quad (\text{A.1})$$

as well as  $\psi_j \rightarrow 0$  as  $j \rightarrow \infty$  and  $\sum_{j=1}^{\infty} j^{1+\vartheta} \psi_j < \infty$  for some  $\vartheta > 0$ .

In order to show (A.1), we write  $x_{tT} \otimes \eta_t = [z'_t \otimes \eta'_t, \varphi(t/T)' \otimes \eta'_t, T^{-1/2}w'_t \otimes \eta'_t]'$  and observe that  $E[z_t \otimes \eta_t] = 0$  and  $E[\eta_t] = 0$ . It follows from the Minkowski inequality that

$$\begin{aligned} \|E_{t-j}(x_{tT} \otimes \eta_t) - E(x_{tT} \otimes \eta_t)\|_2 &\leq \|E_{t-j}(z_t \otimes \eta_t)\|_2 + \|\varphi(t/T) \otimes E_{t-j}(\eta_t)\|_2 \\ &\quad + T^{-1/2} \|E_{t-j}(w_t \otimes \eta_t) - E(w_t \otimes \eta_t)\|_2 \\ &=: A_1 + A_2 + A_3. \end{aligned}$$

<sup>12</sup> Lemma A6 of Bai and Perron (1998) obtains the Hajek-Renyi inequality with the the supremum taken over  $[k_1, \infty]$  rather than the original one with the the supremum taken over a finite range  $[k_1, k_2]$  as in the assertion of this lemma. Their argument, however, can be easily extend to cover the case considered here.

For  $A_1$  and  $A_2$ , an application of the mixing inequality of Ibragimov (1962) yields that<sup>13</sup>

$$A_1 \leq 2(\sqrt{2} + 1)\alpha_j^{1/2-1/\phi} \|z_t \otimes \eta_t\|_\phi \quad \text{and} \quad A_2 \leq 2(\sqrt{2} + 1)\alpha_j^{1/2-1/\phi} \|\eta_t\|_\phi, \quad (\text{A.2})$$

where  $\phi := 4 + \delta$  with  $\delta$  defined in Assumption A2.

For the term  $A_3$ , we separately consider two cases: (i)  $t < j$  and (ii)  $t \geq j$ , given  $t \geq 1$ . First, we consider the case (i), i.e.,  $t - j < 0$ . We have  $w_t = w_0 + \sum_{l=0}^{t-1} u_{w,t-l}$ , which with the Minkowski inequality implies that

$$\sqrt{T}A_3 \leq \|E_{t-j}(w_0 \otimes \eta_t) - E(w_0 \otimes \eta_t)\|_2 + \sum_{l=0}^{t-1} \|E_{t-j}(u_{w,t-l} \otimes \eta_t) - E(u_{w,t-l} \otimes \eta_t)\|_2.$$

Since  $\|E_{t-j}(V) - E(V)\|_2 \leq \|E_{t-j}(V)\|_2$  for a random vector  $V$ , an application of Jensen's inequality and Corollary 14.3 of Davidson (1994)<sup>14</sup> yield that

$$\|E_{t-j}(w_0 \otimes \eta_t) - E(w_0 \otimes \eta_t)\|_2 \leq \|w_0 \otimes \eta_t\|_2 \leq C_3 \alpha_t^{1/2-1/\phi}, \quad (\text{A.3})$$

and that, for  $0 \leq l \leq t-1$

$$\|E_{t-j}(u_{w,t-l} \otimes \eta_t) - E(u_{w,t-l} \otimes \eta_t)\|_2 \leq \|u_{w,t-l} \otimes \eta_t\|_2 \leq C_4 \alpha_l^{1/2-1/\phi}. \quad (\text{A.4})$$

Also, using the mixing inequality of Ibragimov (1962), we can show that

$$\|E_{t-j}(w_0 \otimes \eta_t) - E(w_0 \otimes \eta_t)\|_2 \leq 2(\sqrt{2} + 1)\alpha_{j-t}^{1/2-1/\phi} \|w_0 \otimes \eta_t\|_\phi, \quad (\text{A.5})$$

and that, for  $0 \leq l \leq t-1$ ,

$$\|E_{t-j}(u_{w,t-l} \otimes \eta_t) - E(u_{w,t-l} \otimes \eta_t)\|_2 \leq 2(\sqrt{2} + 1)\alpha_{j-l}^{1/2-1/\phi} \|u_{w,t-l} \otimes \eta_t\|_\phi, \quad (\text{A.6})$$

where both moments on the right-hand side of (A.5) and (A.6) are bounded from Assumption A2. It follows from (A.3)-(A.6) that, when  $t < j$ , we have

$$A_3 \leq C_5 T^{-1/2} \sum_{l=0}^t \min\{\alpha_l^{1/2-1/\phi}, \alpha_{j-l}^{1/2-1/\phi}\} \leq C_5 j^{1/2} \alpha_{[j/2]}^{1/2-1/\phi}, \quad (\text{A.7})$$

where the last inequality is due to that  $\min\{\alpha_l^{1/2-1/\phi}, \alpha_{j-l}^{1/2-1/\phi}\} \leq \alpha_{[j/2]}^{1/2-1/\phi}$  for every  $0 \leq l \leq t$  and that  $T^{-1/2}t \leq t^{1/2} \leq j^{1/2}$  for  $t < j$ .

Next, we consider the case (ii), i.e.,  $0 \leq t - j$ . Since  $w_t = w_{t-j} + \sum_{l=0}^{j-1} u_{w,t-l}$ , the Minkowski inequality leads to that

$$\sqrt{T}A_3 \leq \|w_{t-j} \otimes E_{t-j}(\eta_t)\|_2 + \sum_{l=0}^{j-1} \|E_{t-j}(u_{w,t-l} \otimes \eta_t) - E(u_{w,t-l} \otimes \eta_t)\|_2. \quad (\text{A.8})$$

Using the Cauchy-Schwarz inequality and Ibragimov's mixing inequality, we can show that

$$\|w_{t-j} \otimes E_{t-j}(\eta_t)\|_2 \leq \|w_{t-j}\|_2 \|E_{t-j}(\eta_t)\|_2 \leq \|w_{t-j}\|_2 C_6 \alpha_j^{1/2-1/\phi}. \quad (\text{A.9})$$

Furthermore, we can write  $\|w_{t-j}\|_2^2 = \sum_{s=1}^{t-j} E[u'_{ws} u_{ws}] + 2 \sum_{k=1}^{t-j-1} \sum_{s=1}^{t-j-k} E[u'_{ws} u_{w,s+k}]$ , which with Corollary 14.3 of Davidson (1994) implies

$$T^{-1} \|w_{t-j}\|_2^2 \leq C_7 \left( \frac{t-j}{T} + \sum_{k=1}^{t-j-1} \frac{t-j-k}{T} \alpha_k^{1/2-1/\phi} \right) \leq C_8.$$

<sup>13</sup> For  $A_2$ , we use the fact  $\|\varphi(t/T) \otimes \eta_t\|_2^2 = E[(\varphi(t/T) \otimes \eta_t)'(\varphi(t/T) \otimes \eta_t)] = \varphi(t/T)' \varphi(t/T) E[\eta_t' \eta_t]$ , which implies that  $\|\varphi(t/T) \otimes \eta_t\|_2 \leq C \|\eta_t\|_2$ .

<sup>14</sup> This is a covariance inequality for a  $\alpha$ -mixing sequence. See p. 212 of Davidson (1994) for more details.

Also, applying the same argument used in the case (i), we can show that

$$\sum_{l=0}^{j-1} \|E_{t-j}(u_{w,t-l} \otimes \eta_t) - E(u_{w,t-l} \otimes \eta_t)\|_2 \leq C_9 \sum_{l=0}^{j-1} \min\{\alpha_l^{1/2-1/\phi}, \alpha_{j-l}^{1/2-1/\phi}\}. \quad (\text{A.10})$$

Combining results in (A.9)-(A.10), we obtain

$$A_3 \leq C_{10}(\alpha_j^{1/2-1/\phi} + T^{-1/2} j \alpha_{[j/2]}^{1/2-1/\phi}) \leq C_{11} j^{1/2} \alpha_{[j/2]}^{1/2-1/\phi}.$$

Thus, from the above equation and (A.7), we obtain that  $A_3 \leq C_{12} j^{1/2} \alpha_{[j/2]}^{1/2-1/\phi}$  for every  $t \geq 1$ . This result together with (A.2) and (A.8) yields

$$\|E_{t-j}(x_{tT} \otimes \eta_t) - E(x_{tT} \otimes \eta_t)\|_2 \leq C_{13} j^{1/2} \alpha_{[j/2]}^{1/2-1/\phi}.$$

We set  $\psi_j = j^{1/2} \alpha_{[j/2]}^{1/2-1/\phi}$  and it remains to show that  $\sum_{j=1}^{\infty} j^{1+\vartheta} \psi_j < \infty$  for some  $\vartheta > 0$ . Observe that  $\alpha_{[j/2]}^{1/2-1/\phi} = O(j^{\frac{5}{2}-\frac{1-2\delta}{\delta}})$  under Assumption A1. Thus, for  $\vartheta < (1-2\delta)/\delta$ , we can show that  $\sum_{j=1}^{\infty} j^{1+\vartheta} \psi_j \leq C_{14} \sum_{j=1}^{\infty} j^{-1-\frac{1-2\delta}{\delta}+\vartheta} < \infty$ . Hence we complete the proof. ■

**Lemma A.3.** *For every  $v \in (-1, \infty)$ , we have*

$$-\log(1+v) + \frac{v}{1+v} \leq -\frac{v^2}{2(1+v)(1+|v|)}.$$

**Proof.** Fix  $v \in (-1, \infty)$  and let  $w = -v/(1+v)$ . Then,  $-\log(1+v) = \log(1+w)$ . Since  $\log(1+w) = \int_0^w \frac{1}{1+r} dr$ , we can show that

$$\log(1+w) - w = -\int_0^w \frac{r}{1+r} dr.$$

For  $w > 0$ , we have

$$-\int_0^w \frac{r}{1+r} dr \leq -\frac{1}{1+w} \int_0^w r dr = -\frac{w^2}{2(1+w)},$$

while for  $-1 < w \leq 0$ ,

$$-\int_0^w \frac{r}{1+r} dr = \int_w^0 \frac{r}{1+r} ds \leq \int_w^0 r dr = -\frac{w^2}{2}.$$

Since  $w^2/(1+w) = v^2/(1+v)$  and  $w^2 = v^2/(1+v)^2$ , we can show that,

$$\log(1+w) - w \leq -\frac{v^2}{2(1+v)} \min\left\{1, \frac{1}{1+v}\right\} \leq -\frac{v^2}{2(1+v)(1+|v|)}.$$

Because  $\log(1+w) - w = -\log(1+v) + v/(1+v)$ , the desired result follows. ■

In what follows, we shall use a collection of sub-intervals  $\{[\tau_{l-1}+1, \tau_l]\}_{l=1}^N$  with  $\tau_0 = 0$  and  $\tau_N = T$  as a partition of an interval  $[1, T]$  according to sets of break dates  $\mathcal{K}$  and  $\mathcal{K}^0$ , such that both the true basic parameters and their estimates are constant within each sub-interval and  $N$  is set to be the smallest number of such sub-intervals. That is,  $(\beta_{t,\mathcal{K}}, \beta_{t,\mathcal{K}^0}^0, \Sigma_{t,\mathcal{K}}, \Sigma_{t,\mathcal{K}^0}^0) = (\beta_{\tau_l,\mathcal{K}}, \beta_{\tau_l,\mathcal{K}^0}^0, \Sigma_{\tau_l,\mathcal{K}}, \Sigma_{\tau_l,\mathcal{K}^0}^0)$  for  $\tau_{l-1}+1 \leq t \leq \tau_l$ . For each parameter group  $g \in \{1, \dots, G\}$ , we similarly consider a collection  $\{[\tau_{g,l-1}+1, \tau_{g,l}]\}_{l=1}^{N_g}$  with  $\tau_0 = 0$  and  $\tau_{N_g} = T$  as a partition of an interval  $[1, T]$  given  $\mathcal{K}_g$  and  $\mathcal{K}_g^0$ , where both the true basic parameters and their estimates

of the  $g^{th}$  group are constant within each sub-interval and  $N_g$  is the smallest number of such intervals. Thus we have  $(\beta_{g,t,\mathcal{K}}, \beta_{g,t,\mathcal{K}}^0) = (\beta_{g,\tau_{gl},\mathcal{K}}, \beta_{g,\tau_{gl},\mathcal{K}}^0)$  for  $\tau_{g,l-1} + 1 \leq t \leq \tau_{gl}$  and  $(\Sigma_{t,\mathcal{K}}, \Sigma_{t,\mathcal{K}^0}^0) = (\Sigma_{\tau_{G,l},\mathcal{K}}, \Sigma_{\tau_{G,l},\mathcal{K}^0}^0)$  for  $\tau_{G,l-1} + 1 \leq t \leq \tau_{G,l}$ , whereas the basic parameters of the other groups may change. For  $\tau_{G,l-1} + 1 \leq t \leq \tau_{G,l}$  with  $l \in \{1, \dots, N_g\}$ , we define

$$\Psi_l := (\Sigma_{t,\mathcal{K}^0}^0)^{-1/2}(\Sigma_{t,\mathcal{K}} - \Sigma_{t,\mathcal{K}^0}^0)(\Sigma_{t,\mathcal{K}^0}^0)^{-1/2}, \quad (\text{A.11})$$

where we have  $I_n + \Psi_l = (\Sigma_{\tau_{G,l},\mathcal{K}^0}^0)^{-1/2}\Sigma_{\tau_{G,l},\mathcal{K}}(\Sigma_{\tau_{G,l},\mathcal{K}^0}^0)^{-1/2}$ . Since  $\Psi_l$  is an  $n \times n$  symmetric matrix, there exists an orthogonal matrix  $U$  such that

$$U\Psi U' = \text{diag}\{\lambda_{l1}^\Psi, \dots, \lambda_{ln}^\Psi\} \quad \text{and} \quad U(I_n + \Psi)U' = \text{diag}\{1 + \lambda_{l1}^\Psi, \dots, 1 + \lambda_{ln}^\Psi\},$$

where  $\lambda_{l1}^\Psi, \dots, \lambda_{ln}^\Psi$  are the eigenvalues of  $\Psi_l$ .

In the below lemma, we shall obtain an upper bound for the normalized log likelihood based on sub-intervals. As short-hand notation, we define, for  $1 \leq t \leq T$  and  $1 \leq g \leq G$ ,

$$\Delta\beta_{t,\mathcal{K}} := \beta_{t,\mathcal{K}} - \beta_{t,\mathcal{K}^0}^0 \quad \text{and} \quad \Delta\beta_{g,t,\mathcal{K}} := \beta_{g,t,\mathcal{K}} - \beta_{g,t,\mathcal{K}^0}^0.$$

**Lemma A.4.** *Suppose that Assumption A1-A5 hold. Then,*

$$\ell_T(\mathcal{K}, \theta) \leq C \left\{ \sum_{g=1}^G \sum_{l=1}^{N_g} \bar{\ell}_{g,l}(\mathcal{K}, \theta) + \sum_{l=1}^{N_G} \bar{\ell}_{G+1,l}(\mathcal{K}, \theta) + \Delta_T(\mathcal{K}, \theta) \right\},$$

where, for  $g = 1, \dots, G$ ,

$$\begin{aligned} \bar{\ell}_{g,l}(\mathcal{K}, \theta) &:= \left( \left\| \sum_{t=\tau_{g,l-1}+1}^{\tau_{gl}} X_{tT} \Sigma_{t,\mathcal{K}}^{-1} u_t \right\| - (\tau_{gl} - \tau_{g,l-1}) \|\Delta\beta_{g,\tau_{gl},\mathcal{K}}\| \right) \|\Delta\beta_{g,\tau_{gl},\mathcal{K}}\|, \\ \bar{\ell}_{G+1,l}(\mathcal{K}, \theta) &:= \sum_{i=1}^n \left( \left\| \sum_{t=\tau_{G,l-1}+1}^{\tau_{Gl}} (\eta_t \eta_t' - I_n) \right\| - (\tau_{Gl} - \tau_{G,l-1}) |\lambda_{il}^\Psi| \right) |\lambda_{il}^\Psi|, \\ \Delta_T(\mathcal{K}, \theta) &:= \max_{1 \leq t \leq T} \|\Delta\beta_{t,\mathcal{K}}\|. \end{aligned}$$

**Proof.** We can write  $\log f(y_t | X_{tT}, \theta_{t,\mathcal{K}}) = -(1/2) (\log(2\pi)^n + \log |\Sigma_{t,\mathcal{K}}| + \|\Sigma_{t,\mathcal{K}}^{-1/2}(u_t - X_{tT}' \Delta\beta_{t,\mathcal{K}})\|^2)$ , which implies that

$$\begin{aligned} \ell_T(\mathcal{K}, \theta) &= -\frac{1}{2} \sum_{t=1}^T (\log |\Sigma_{t,\mathcal{K}}| - \log |\Sigma_{t,\mathcal{K}^0}^0| + \|\Sigma_{t,\mathcal{K}}^{-1/2} u_t\|^2 - \|(\Sigma_{t,\mathcal{K}^0}^0)^{-1/2} u_t\|^2) \\ &\quad + \sum_{t=1}^T \Delta\beta_{t,\mathcal{K}}' X_{tT} \Sigma_{t,\mathcal{K}}^{-1} u_t - \frac{1}{2} \sum_{t=1}^T \|\Sigma_{t,\mathcal{K}}^{-1/2} X_{tT}' \Delta\beta_{t,\mathcal{K}}\|^2 \\ &=: A_1 + A_2 + A_3. \end{aligned}$$

For the term  $A_1$ , we write  $\log |\Sigma_{t,\mathcal{K}}| - \log |\Sigma_{t,\mathcal{K}^0}^0| = \log |(\Sigma_{t,\mathcal{K}^0}^0)^{-1/2} \Sigma_{t,\mathcal{K}} (\Sigma_{t,\mathcal{K}^0}^0)^{-1/2}|$  and also  $u_t = (\Sigma_{t,\mathcal{K}^0}^0)^{1/2} \eta_t$ . Since the term  $A_1$  depends only on  $\mathcal{K}_G$  and  $\mathcal{K}_G^0$ , we can express

$$A_1 = \sum_{l=1}^{N_G} \left\{ -\frac{1}{2} \sum_{t=\tau_{G,l-1}+1}^{\tau_{Gl}} \left( \log |I_n + \Psi_l| + \text{tr}((I_n + \Psi_l)^{-1} \eta_t \eta_t') - \text{tr}(\eta_t \eta_t') \right) \right\} =: \sum_{l=1}^{N_G} A_{1,l}.$$

For every  $l = 1, \dots, N_G$ , we have that  $\log |I_n + \Psi_l| = \sum_{i=1}^n \log(1 + \lambda_{li}^\Psi)$  and that

$$\text{tr}((I_n + \Psi_l)^{-1} \eta_t \eta_t') = \text{tr} \left( \text{diag} \left( \left\{ \frac{1}{1 + \lambda_{li}^\Psi} \right\}_{i=1}^n \right) U' \eta_t \eta_t' U \right),$$

which leads to

$$A_{1,l} = -\frac{\tau_{Gl} - \tau_{G,l-1}}{2} \sum_{i=1}^n \log(1 + \lambda_{li}^\Psi) + \frac{1}{2} \text{tr} \left( \text{diag} \left( \left\{ \frac{\lambda_{li}^\Psi}{1 + \lambda_{li}^\Psi} \right\}_{i=1}^n \right) U' \sum_{t=\tau_{G,l-1}+1}^{\tau_{Gl}} \eta_t \eta_t' U \right).$$

Lemma A.3 implies that  $-\log(1 + \lambda_{li}^\Psi) + \lambda_{li}^\Psi / (1 + \lambda_{li}^\Psi) \leq -(\lambda_{li}^\Psi)^2 / 2(1 + \lambda_{li}^\Psi)(1 + |\lambda_{li}^\Psi|)$ . Thus,

$$\begin{aligned} A_{1,l} \leq & -\frac{\tau_{Gl} - \tau_{G,l-1}}{2} \sum_{i=1}^n \frac{|\lambda_{li}^\Psi|^2}{2(1 + \lambda_{li}^\Psi)(1 + |\lambda_{li}^\Psi|)} \\ & + \frac{1}{2} \text{tr} \left( \text{diag} \left( \left\{ \frac{\lambda_{li}^\Psi}{1 + \lambda_{li}^\Psi} \right\}_{i=1}^n \right) U' \sum_{t=\tau_{G,l-1}+1}^{\tau_{Gl}} (\eta_t \eta_t' - I_n) U \right). \end{aligned}$$

Since the maximum of diagonal elements of  $U' \sum_{t=\tau_{G,l-1}+1}^{\tau_{Gl}} (\eta_t \eta_t' - I_n) U$  is bounded from above by  $\|U' \sum_{t=\tau_{G,l-1}+1}^{\tau_{Gl}} (\eta_t \eta_t' - I_n) U\|$  with  $\|U\| = 1$ ,

$$A_{1,l} \leq \frac{1}{2} \sum_{i=1}^n \left\{ -\frac{(\tau_{Gl} - \tau_{G,l-1})|\lambda_{li}^\Psi|^2}{2(1 + \lambda_{li}^\Psi)(1 + |\lambda_{li}^\Psi|)} + \frac{|\lambda_{li}^\Psi|}{1 + \lambda_{li}^\Psi} \left\| \sum_{t=\tau_{G,l-1}+1}^{\tau_{Gl}} (\eta_t \eta_t' - I_n) \right\| \right\}. \quad (\text{A.12})$$

From the compactness of  $\Theta$  and (A.11), we have that  $\max_{1 \leq i \leq n} (1 + \lambda_{li}^\Psi) = \|I_n + \Psi_l\| \leq C_1$  and also that

$$1 + \min_{1 \leq i \leq n} \lambda_{li}^\Psi = \min_{a \in \mathbb{R}^n} \frac{a'(I_n + \Psi_l)a}{a'a} \geq \min_{b \in \mathbb{R}^n} \frac{b' \Sigma_{\tau_{Gl}, \mathcal{K}} b}{b'b} \cdot \min_{a \in \mathbb{R}^n} \frac{a'(\Sigma_{\tau_{Gl}, \mathcal{K}^0}^0)^{-1} a}{a'a} \geq C_2$$

Thus we have that, for every  $i = 1, \dots, n$ ,

$$\frac{1}{C_1} \leq \frac{1}{1 + \lambda_{li}^\Psi} \leq \frac{1}{C_2} \quad \text{and} \quad -\frac{1}{1 + |\lambda_{li}^\Psi|} \leq -\frac{1}{1 + \max_{1 \leq j \leq 2} |C_j - 1|}.$$

This together with (A.12) yields

$$A_{1,l} \leq C_3 \sum_{i=1}^n \left\{ -(\tau_{Gl} - \tau_{G,l-1})|\lambda_{li}^\Psi|^2 + |\lambda_{li}^\Psi| \left\| \sum_{t=\tau_{G,l-1}+1}^{\tau_{Gl}} (\eta_t \eta_t' - I_n) \right\| \right\}.$$

It follows that  $A_1 \leq C_4 \sum_{l=1}^{N_G} \bar{\ell}_{G+1,l}(\mathcal{K}, \theta)$ .

We now consider  $A_2$  and  $A_3$ . Observe that  $\Delta\beta_{t,\mathcal{K}} = \sum_{g=1}^G \Delta\beta_{g,t,\mathcal{K}}$ . We have

$$A_2 = \sum_{g=1}^G \sum_{t=1}^T \Delta\beta'_{g,t,\mathcal{K}} X_{tT} \Sigma_{t,\mathcal{K}}^{-1} u_t. \quad (\text{A.13})$$

Also, given that  $X_{tT} \Sigma_{t,\mathcal{K}}^{-1} X_{tT}' = S'(x_{tT} x_{tT}' \otimes \Sigma_{\tau_l, \mathcal{K}}^{-1}) S$  for  $\tau_{l-1} + 1 \leq t \leq \tau_l$ , we can show that

$$A_3 = \sum_{l=1}^N \left\{ -\frac{1}{2} \left\| \left( \sum_{t=\tau_{l-1}+1}^{\tau_l} x_{tT} x_{tT}' \otimes \Sigma_{\tau_l, \mathcal{K}}^{-1} \right)^{1/2} S \Delta\beta_{\tau_l, \mathcal{K}} \right\|^2 \right\} =: \sum_{l=1}^N A_{3,l}.$$

Under Assumption A2, there exists a finite integer  $k_0$  such that the minimum eigenvalue of  $(\tau_l - \tau_{l-1})^{-1} \sum_{t=\tau_{l-1}+1}^{\tau_l} x_{tT} x_{tT}'$  is strictly positive for every  $(\tau_l - \tau_{l-1}) \geq k_0$  and also eigenvalues of  $\Sigma_{\tau_l, \mathcal{K}}$  are finite positive values in  $\Theta$ . Thus, an application of the result that  $\min_{1 \leq i \leq n} \lambda_i(A) \|b\|^2 \leq b' A b \leq \max_{1 \leq i \leq n} \lambda_i(A) \|b\|^2$  for an  $n \times 1$  vector  $b$  and an  $n \times n$  symmetric

matrix  $A$  with eigenvalues  $\{\lambda_i(A)\}_{i=1}^n$  yields that, when  $\tau_l - \tau_{l-1} \geq k_0$ ,

$$A_{3,l} \leq -C_5(\tau_l - \tau_{l-1})\|S\Delta\beta_{\tau_l, \mathcal{K}}\|^2 \leq -C_6(\tau_l - \tau_{l-1})\|\Delta\beta_{\tau_l, \mathcal{K}}\|^2, \quad (\text{A.14})$$

where the last inequality is due to that  $S'S$  is positive definite.<sup>15</sup> When  $\tau_l - \tau_{l-1} < k_0$ , we have that  $(\tau_l - \tau_{l-1})\|\Delta\beta_{\tau_l, \mathcal{K}}\|^2 \leq C_7\|\Delta\beta_{\tau_l, \mathcal{K}}\|^2$ , which yields

$$A_{3,l} \leq 0 \leq -C_8(\tau_l - \tau_{l-1})\|\Delta\beta_{\tau_l, \mathcal{K}}\|^2 + C_9\|\Delta\beta_{\tau_l, \mathcal{K}}\|^2. \quad (\text{A.15})$$

It follows from (A.14) and (A.15) that  $A_3 \leq -C_{10} \sum_{l=1}^N (\tau_l - \tau_{l-1}) \|\beta_{\tau_l, \mathcal{K}} - \beta_{\tau_l, \mathcal{K}^0}^0\|^2 + C_{11} \Delta_T(\mathcal{K}, \theta)$ . Furthermore, we can show that  $\sum_{l=1}^N (\tau_l - \tau_{l-1}) \|\Delta\beta_{\tau_l, \mathcal{K}}\|^2 = \sum_{t=1}^T \|\Delta\beta_{t, \mathcal{K}}\|^2$  and also that  $\|\Delta\beta_{t, \mathcal{K}}\|^2 = \sum_{g=1}^G \|\Delta\beta_{g,t, \mathcal{K}}\|^2$  because  $(\Delta\beta_{g_1, t, \mathcal{K}})' \Delta\beta_{g_2, t, \mathcal{K}} = 0$  for all  $g_1, g_2 \in \{1, \dots, G\}$  with  $g_1 \neq g_2$ . Thus we have

$$A_3 \leq -C_{12} \sum_{g=1}^G \sum_{t=1}^T \|\Delta\beta_{g,t, \mathcal{K}}\|^2 + C_{13} \Delta_T(\mathcal{K}, \theta). \quad (\text{A.16})$$

For each  $g = 1, \dots, G$ , we have partitions  $\{[\tau_{g,l-1} + 1, \tau_{gl}]\}$  of an interval  $[1, T]$ . It follows from (A.13) and (A.16) that  $A_2 + A_3 \leq C_{14} \{\sum_{g=1}^G \sum_{l=1}^{N_g} \bar{\ell}_{g,l}(\mathcal{K}, \theta) + \Delta_T(\mathcal{K}, \theta)\}$ . Hence we obtain the desired result.  $\blacksquare$

We shall established several properties of the terms  $\{\bar{\ell}_{g,l}(\mathcal{K}, \theta)\}_{g=1}^{G+1}$  based on subsample free from structural changes. To this end, we consider a sequence  $\{\xi_t\}_{t=1}^T$  of some random vectors or matrices satisfying the condition under which the Hajek-Renyi inequality in Lemma A.2 holds. Let  $\gamma$  be a parameter vector or matrices as an element of the bounded parameter space  $\Gamma := \{\gamma : \|\gamma\| \leq C\}$ . We define an object depending on a subsample of  $k$  observations free from structural changes in  $\gamma$ : for  $k = 1, \dots, T$ ,

$$\ell_k^{(0)}(\gamma) := \left( \left\| \sum_{t=1}^k \xi_t \right\| - k\|\gamma\| \right) \|\gamma\|.$$

We now establish a series of properties related to the likelihood function that will enable us to prove the rate of convergence of the estimates.

### Property 1.

$$\sup_{1 \leq k \leq T} \sup_{\gamma \in \Gamma} \ell_k^{(0)}(\gamma) \leq |O_p(\log T)|.$$

**Proof.** Let  $D > 0$  and define  $\Gamma_{1,k}(D) := \{\gamma \in \mathcal{G} : \sqrt{k}\|\gamma\| \leq D(\log T)^{1/2}\}$  for  $1 \leq k \leq T$ . We can write  $\ell_k^{(0)}(\gamma) = (k^{-1/2} \|\sum_{t=1}^k \xi_t\| - \sqrt{k}\|\gamma\|) \sqrt{k}\|\gamma\|$  for every  $1 \leq k \leq T$ . It follows that, for any  $1 \leq k \leq T$ ,

$$\sup_{\gamma \in \Gamma \setminus \Gamma_{1,k}(D)} \ell_k^{(0)}(\gamma) \leq \sup_{\gamma \in \Gamma \setminus \Gamma_{1,k}(D)} \left( \frac{1}{\sqrt{k}} \left\| \sum_{t=1}^k \xi_t \right\| - D(\log T)^{1/2} \right) \sqrt{k}\|\gamma\|,$$

and

$$\sup_{\gamma \in \Gamma_{1,k}(D)} \ell_k^{(0)}(\gamma) \leq \frac{1}{\sqrt{k}} \left\| \sum_{t=1}^k \xi_t \right\| D(\log T)^{1/2}.$$

<sup>15</sup> The selection matrix  $S$  is of dimension  $nq \times p$  with full column rank and thus  $Sv \neq 0$  for all  $v \in \mathbb{R}^p$  with  $v \neq 0$ . It follows that  $v'S'Sv \neq 0$  for all  $v \in \mathbb{R}^p$  with  $v \neq 0$  and  $S'S$  is positive definite. This implies that there exists a constant  $c > 0$  such that  $\|Sb\| \geq c\|b\|$  for any  $b \in \mathbb{R}^p$ .

Lemma A.2 implies that, for any  $B_1 > 0$ ,

$$\Pr \left\{ \sup_{1 \leq k \leq T} \frac{1}{\sqrt{k \log T}} \left\| \sum_{t=1}^k \xi_t \right\| \geq B_1 \right\} \leq \frac{C_1}{B_1^2 \log T} \sum_{k=1}^T \frac{1}{k}.$$

The right-hand side of the above inequality becomes arbitrarily small for a sufficiently large  $B_1$  because  $\sum_{k=1}^T k^{-1} = O(\log T)$ . Thus,  $\sup_{1 \leq k \leq T} k^{-1/2} \left\| \sum_{t=1}^k \xi_t \right\| - D(\log T)^{1/2} < 0$  in probability, for a sufficiently large  $D$ . It follows that, for a sufficiently large  $D > 0$ ,

$$\sup_{1 \leq k \leq T} \sup_{\gamma \in \Gamma \setminus \Gamma_{1,k}(D)} \ell_k^{(0)}(\gamma) \leq -C_2 D^2 \log T \quad \text{and} \quad \sup_{1 \leq k \leq T} \sup_{\gamma \in \Gamma_{1,k}(D)} \ell_k^{(0)}(\gamma) \leq C_3 D \log T,$$

in probability. Hence, the desired conclusion follows.  $\blacksquare$

**Property 2.** For any  $D > 0$ , there exists a constant  $A > 0$  such that, for any deterministic sequence  $m_T \geq A v_T^{-2}$ ,

$$\sup_{m_T \leq k \leq T} \sup_{\gamma: \|\gamma\| \geq D v_T} \ell_k^{(0)}(\gamma) \leq -|O_p((D v_T)^2 m_T)|.$$

**Proof.** Let  $D > 0$  be fixed. We have, for every  $1 \leq k \leq T$ ,

$$\sup_{\gamma: \|\gamma\| \geq D v_T} \frac{1}{k} \ell_k^{(0)}(\gamma) \leq \sup_{\gamma: \|\gamma\| \geq D v_T} \left( \frac{1}{k} \left\| \sum_{t=1}^k \xi_t \right\| - D v_T \right) \|\gamma\|.$$

Lemma A.2 yields that, for any  $A > 0$  and for any  $\epsilon > 0$ ,

$$\Pr \left\{ \sup_{A v_T^{-2} \leq k \leq T} \frac{1}{k v_T} \left\| \sum_{t=1}^k \xi_{tT} \right\| > \epsilon \right\} \leq \frac{C_1}{\epsilon^2} \left( \frac{1}{A} + \frac{1}{v_T^2} \sum_{k=A v_T^{-2}}^T \frac{1}{k^2} \right). \quad (\text{A.17})$$

Because  $\sum_{k=A v_T^{-2}}^T k^{-2} = O((A v_T^{-2})^{-1})$ , we can show that the right-hand side of (A.17) becomes arbitrarily small for a sufficiently large  $A > 0$ . Since  $\epsilon$  can be arbitrarily small, there exists an  $A$  such that

$$\sup_{A v_T^{-2} \leq k \leq T} \sup_{\gamma: \|\gamma\| \geq D v_T} \frac{1}{k} \ell_k^{(0)}(\gamma) \leq -C_2 (D v_T)^2.$$

in probability. The desired result follows because  $-m_T^{-1} \leq -k^{-1}$  when  $k \geq m_T$ .  $\blacksquare$

**Property 3.** Let  $\Gamma_3(D) := \{\gamma \in \mathcal{G} : \sqrt{T} \|\gamma\| \leq D\}$  for any  $D > 0$ . Then, for any  $\delta \in (0, 1)$ , (a) there exists a  $D > 0$  such that

$$\sup_{\delta T \leq k \leq T} \sup_{\gamma \in \Gamma \setminus \Gamma_3(D)} \ell_k^{(0)}(\gamma) \leq -|O_p(D^2)|,$$

(b) for any  $D > 0$ ,

$$\sup_{\delta T \leq k \leq T} \sup_{\gamma \in \Gamma_3(D)} \ell_k^{(0)}(\gamma) = O_p(D).$$

**Proof.** Let  $\delta \in (0, 1)$  be fixed. Then, we have, for every  $\delta T \leq k \leq T$  and for any  $D > 0$ ,

$$\sup_{\gamma \in \Gamma \setminus \Gamma_3(D)} \ell_k^{(0)}(\gamma) \leq \sup_{\gamma \in \Gamma \setminus \Gamma_3(D)} \left( \frac{1}{\sqrt{T}} \left\| \sum_{t=1}^k \xi_t \right\| - \delta D \right) \sqrt{T} \|\gamma\|, \quad (\text{A.18})$$



and

$$\sup_{\gamma \in \Gamma_3(D)} |\ell_k^{(0)}(\gamma)| \leq \frac{1}{\sqrt{T}} \left\| \sum_{t=1}^k \xi_t \right\| D. \quad (\text{A.19})$$

Lemma A.2 implies that  $\sup_{\delta T \leq k \leq T} \left\| \sum_{t=1}^k \xi_t \right\| = O_p(\sqrt{T})$ . It follows from (A.18) that, for some  $D > 0$ ,  $\sup_{\delta T \leq k \leq T} \sup_{\gamma \in \Gamma \setminus \Gamma_3(D)} \ell_k^{(0)}(\gamma) \leq -C_1 D^2$  in probability, while it follows from (A.19) that  $\sup_{\delta T \leq k \leq T} \sup_{\gamma \in \Gamma_3(D)} |\ell_k^{(0)}(\gamma)| \leq C_2 D$  in probability, for any  $D > 0$ . Hence, the desired result follows.  $\blacksquare$

**Property 4.** For any constant  $M > 0$  and a deterministic sequence  $b_T > 0$ , we have

$$\sup_{1 \leq k \leq M v_T^{-2}} \sup_{\gamma: \|\gamma\| \leq b_T} \ell_k^{(0)}(\gamma) = O_p(M^{1/2} v_T^{-1} b_T).$$

**Proof.** Let  $M > 0$ . Then,  $\sup_{1 \leq k \leq M v_T^{-2}} \sup_{\gamma: \|\gamma\| \leq b_T} |\ell_k^{(0)}(\gamma)| \leq \sup_{1 \leq k \leq M v_T^{-2}} \left\| \sum_{t=1}^k \xi_t \right\| b_T$ . Lemma A.2 yields that  $\sup_{1 \leq k \leq M v_T^{-2}} \left\| \sum_{t=1}^k \xi_t \right\| \leq O_p((M v_T^{-2})^{1/2})$ , completes the proof.  $\blacksquare$

For  $\tau_{G,l-1} + 1 \leq t \leq \tau_{Gl}$ , we can show that

$$\|\Psi_l\| \leq \|(\Sigma_{t,\mathcal{K}^0}^0)^{-1/2}\|^2 \|\Sigma_{t,\mathcal{K}} - \Sigma_{t,\mathcal{K}^0}^0\| \quad \text{and} \quad \|\Sigma_{t,\mathcal{K}} - \Sigma_{t,\mathcal{K}^0}^0\| \leq \|(\Sigma_{t,\mathcal{K}^0}^0)^{1/2}\|^2 \cdot \|\Psi_l\|,$$

Since  $\|(\Sigma_{t,\mathcal{K}^0}^0)^{1/2}\|$  and  $\|(\Sigma_{t,\mathcal{K}^0}^0)^{-1/2}\|$  are bounded and  $\|\Psi_l\| = \max_{1 \leq i \leq n} |\lambda_{il}^\Psi|$ , we have

$$d_1 \|\Sigma_{t,\mathcal{K}} - \Sigma_{t,\mathcal{K}^0}^0\| \leq \max_{1 \leq i \leq n} |\lambda_{il}^\Psi| \leq d_2 \|\Sigma_{t,\mathcal{K}} - \Sigma_{t,\mathcal{K}^0}^0\|,$$

for some positive constants  $d_1$  and  $d_2$ . This relation will be used when we restricts a space for the covariance matrix of the error. In the proposition below, we will establish the result regarding the break date estimates.

**Proposition A.1.** Under Assumption A1-A5, there exists a  $B > 0$  such that

$$\lim_{T \rightarrow \infty} \Pr \left\{ |\hat{k}_{gj} - k_{gj}^0| > B v_T^{-2} \log T \right\} = 0,$$

for every  $(g, j) \in \{1, \dots, G\} \times \{1, \dots, m\}$ .

**Proof.** For a constant  $B > 0$ , define

$$\ddot{\Xi}(B) := \left\{ \mathcal{K} \in \Xi_\nu : \max_{1 \leq g \leq G} \max_{1 \leq j \leq m} |k_{gj} - k_{gj}^0| \leq B v_T^{-2} \log T \right\}.$$

To prove the assertion, we shall show that, for a sufficiently large  $B > 0$ ,

$$\lim_{T \rightarrow \infty} \Pr \left\{ \sup_{(\mathcal{K}, \theta) \in \Xi_\nu \setminus \ddot{\Xi}(B) \times \Theta} \ell_T(\mathcal{K}, \theta) \geq 0 \right\} = 0. \quad (\text{A.20})$$

Since the normalized log likelihood evaluated at the maximum likelihood estimates should be non-negative, the desired conclusion follows from (A.20).

For this purpose, we examine the upper bound in Lemma A.4 given sets of break dates  $\mathcal{K} \notin \ddot{\Xi}(B)$  and  $\mathcal{K}^0$ . First, observe that Property 1 provides a not necessarily sharp but general upper bound in probability and that the parameter space is bounded. Thus,

$$\sup_{(\mathcal{K}, \theta) \in \Xi_\nu \setminus \ddot{\Xi}(B) \times \Theta} \bar{\ell}_{g,l}(\mathcal{K}, \theta) \leq |O_P(\log T)| \quad \text{and} \quad \sup_{(\mathcal{K}, \theta) \in \Xi_\nu \setminus \ddot{\Xi}(B) \times \Theta} \Delta(\mathcal{K}, \theta) \leq C_1, \quad (\text{A.21})$$

for every  $1 \leq g \leq G+1$  and  $1 \leq l \leq 2(m+1)$ .

Next, for  $\mathcal{K} \notin \ddot{\Xi}(B)$ , there exists a pair  $(g, j) \in \{1, \dots, G\} \times \{1, \dots, m\}$  such that some neighborhood  $\mathcal{N}_{gj} := \{t \in [1, T] : |t - k_{gj}^0| \leq Bv_T^{-2} \log T\}$  of a true break date  $k_{gj}^0$  contains none of break dates  $\mathcal{K}_g$  of the  $g^{th}$  group, i.e.,  $\mathcal{K}_g \not\subset \mathcal{N}_{gj}$ . This implies that there is  $\tau_{gl} = k_{gj}^0$  with a union of sub-intervals

$$[\tau_{g,l-1}+1, \tau_{gl}] \cup [\tau_{gl}+1, \tau_{g,l+1}] \quad \text{with} \quad \min_{l \leq j \leq l+1} (\tau_{gj} - \tau_{g,j-1}) \geq Bv_T^{-2} \log T.$$

Since  $\mathcal{K}_g \not\subset (\tau_{g,l-1}, \tau_{g,l+1})$ , the  $g^{th}$  group estimates are constant for  $\tau_{g,l-1}+1 \leq t \leq \tau_{g,l+1}$  and both terms  $\bar{\ell}_{g,l}(\mathcal{K}, \theta)$  and  $\bar{\ell}_{g,l+1}(\mathcal{K}, \theta)$  depend on the same  $g^{th}$  group estimates. Thus note that the triangle inequality yields that

$$C_2 v_T \leq 2 \max \left\{ \|\beta_{g,\tau_{g,l+1},\mathcal{K}} - \beta_{g,\tau_{gl},\mathcal{K}^0}^0\|, \|\beta_{g,\tau_{g,l+1},\mathcal{K}} - \beta_{g,\tau_{g,l+1},\mathcal{K}^0}^0\| \right\},$$

and additionally when  $g = G$ ,

$$C_3 v_T \leq 2 \max \left\{ \|\Sigma_{\tau_{G,l+1},\mathcal{K}} - \Sigma_{\tau_{Gl},\mathcal{K}^0}^0\|, \|\Sigma_{\tau_{G,l+1},\mathcal{K}} - \Sigma_{\tau_{G,l+1},\mathcal{K}^0}^0\| \right\}.$$

This implies that either one of  $\bar{\ell}_{g,l}(\mathcal{K}, \theta)$  and  $\bar{\ell}_{g,l+1}(\mathcal{K}, \theta)$  satisfies the condition in Property 2 with  $m_T = Bv_T^{-2} \log T$ , which together with (A.21) implies that, for a sufficiently large  $B$ ,

$$\sup_{(\mathcal{K}, \theta) \in \Xi_\nu \setminus \ddot{\Xi}(B) \times \Theta} \ell_T(\mathcal{K}, \theta) \leq -|O_p(B \log T)| + O_p(\log T).$$

This yields (A.20) for a sufficiently large  $B$  and thus we complete the proof.  $\blacksquare$

**Proposition A.2.** *Suppose that Assumption A1-A5 hold. Then,*

$$\hat{\beta}_{gj} - \beta_{gj}^0 = o_p(v_T) \quad \text{and} \quad \hat{\Sigma}_j - \Sigma_j^0 = o_p(v_T),$$

for every  $(g, j) \in \{1, \dots, G\} \times \{1, \dots, m+1\}$ .

**Proof.** Let  $\epsilon > 0$  be fixed and define a subset of the parameter space  $\Theta$ :

$$\ddot{\Theta}(\epsilon) := \left\{ \theta \in \Theta : \max_{1 \leq g \leq G} \max_{1 \leq j \leq m+1} \|\beta_{gj} - \beta_{gj}^0\| \leq \epsilon v_T \text{ and } \max_{1 \leq j \leq m+1} \|\Sigma_j - \Sigma_j^0\| \leq \epsilon v_T \right\}.$$

Proposition A.1 shows that the break date estimates  $\hat{\mathcal{K}}$  are included in  $\ddot{\Xi}(B)$  in probability for a sufficiently large  $B$  and thus we consider the case where  $\mathcal{K} \in \ddot{\Xi}(B)$ . For  $\theta \in \Theta \setminus \ddot{\Theta}(\epsilon)$ , there exists a pair  $(g, j) \in \{1, \dots, G\} \times \{1, \dots, m\}$  such that either

$$\|\beta_{gj} - \beta_{gj}^0\| \geq \epsilon v_T \quad \text{or} \quad \|\Sigma_j - \Sigma_j^0\| \geq \epsilon v_T. \quad (\text{A.22})$$

Observe that  $k_{gj} - k_{g,j-1} \geq \nu T$  and  $k_{gj}^0 - k_{g,j-1}^0 \geq \nu T$ , while  $|k_{gj} - k_{gj}^0| \leq Bv_T^{-2} \log T$ . For some  $l \in \{1, \dots, N_g\}$ , we have  $\tau_{g,l-1} = \max\{k_{g,j-1}, k_{g,j-1}^0\}$  and  $\tau_{gl} = \max\{k_{g,j}, k_{gj}^0\}$  satisfying that  $\tau_{gl} - \tau_{g,l-1} \geq \delta T$  for some  $\delta \in (0, 1)$  and that (A.22) holds over a sub-interval  $[\tau_{g,l-1}+1, \tau_{gl}]$ . Thus, Property 2 with  $m_T = \delta T$  implies that

$$\sup_{(\mathcal{K}, \theta) \in \ddot{\Xi}(B) \times \Theta \setminus \ddot{\Theta}(\epsilon)} \bar{\ell}_{g,l}(\mathcal{K}, \theta) \leq -|O_p(\epsilon^2 T v_T^2)|.$$

For the other sub-intervals, Property 1 provides an upper bound  $|O_p(\log T)|$ . Since  $\sqrt{T} v_T / \log T \rightarrow \infty$  as  $T \rightarrow \infty$ , we can show that

$$\sup_{(\mathcal{K}, \theta) \in \ddot{\Xi}(B) \times \Theta \setminus \ddot{\Theta}(\epsilon)} \ell_T(\mathcal{K}, \theta) \leq -|O_p(\epsilon^2 T v_T^2)|.$$

This leads to the desired result.  $\blacksquare$

**Proof of Theorem 1.** (a) Proposition A.1 shows that  $\hat{\mathcal{K}} \in \ddot{\Xi}(B)$  in probability for some  $B > 0$ , while both  $\hat{\mathcal{K}}$  and  $\mathcal{K}^0$  are included in  $\Xi_\nu$ . Thus, it suffices to consider the case where either  $\tau_{gl} - \tau_{g,l-1} \geq \delta T$  for some  $\delta > 0$  or  $\tau_{gl} - \tau_{g,l-1} \leq Bv_T^{-2} \log T$  for every  $(g, l) \in \{1, \dots, G\} \times \{1, \dots, N\}$ . If  $\tau_{gl} - \tau_{g,l-1} \geq \delta T$ , then Property 3(a) and (b) imply that

$$\sup_{(\mathcal{K}, \theta) \in \ddot{\Xi}(B) \times \Theta} \bar{\ell}_{g,l}(\mathcal{K}, \theta) \leq |O_p(1)|. \quad (\text{A.23})$$

When  $\tau_{gl} - \tau_{g,l-1} \leq Bv_T^{-2} \log T$ , there are two cases:  $Mv_T^{-2} \leq \tau_{gl} - \tau_{g,l-1} \leq Bv_T^{-2} \log T$  and  $\tau_{gl} - \tau_{g,l-1} \leq Mv_T^{-2}$  for some  $M > 0$ . For the sake of concreteness, let  $\tau_{g,l-1} = k_{gj}^0$  and  $\tau_{gl} = \hat{k}_{gj}$  in both cases. When  $k_{gj}^0 + 1 \leq t \leq \hat{k}_{gj}$ , we have  $(\hat{\beta}_{g,t,\hat{\mathcal{K}}}, \beta_{g,t,\mathcal{K}^0}^0) = (\hat{\beta}_{gj}, \beta_{g,j+1}^0)$  for  $1 \leq g \leq G$  and  $(\hat{\Sigma}_{t,\hat{\mathcal{K}}}, \Sigma_{t,\mathcal{K}^0}^0) = (\hat{\Sigma}_j, \Sigma_{j+1}^0)$  for  $g = G$ . Since  $\|\beta_{g,j+1}^0 - \beta_{gj}^0\| = v_T \|\delta_{gj}\|$  and  $\|\Sigma_{j+1}^0 - \Sigma_j^0\| = v_T \|\Phi_j\|$ , we can show<sup>16</sup>

$$\left| \|\hat{\beta}_{gj} - \beta_{g,j+1}^0\| - v_T \|\delta_{gj}\| \right| \leq \|\hat{\beta}_{gj} - \beta_{gj}^0\| \quad \text{and} \quad \left| \|\hat{\Sigma}_j - \Sigma_{j+1}^0\| - v_T \|\Phi_j\| \right| \leq \|\hat{\Sigma}_j - \Sigma_j^0\|.$$

Moreover, Proposition A.2 shows that  $\|\hat{\beta}_{gj} - \beta_{gj}^0\| = o_p(v_T)$  and  $\|\hat{\Sigma}_j - \Sigma_j^0\| = o_p(v_T)$ . Thus,

$$\|\hat{\beta}_{gj} - \beta_{g,j+1}^0\| = v_T \|\delta_{gj}\| + o_p(v_T) \quad \text{and} \quad \|\hat{\Sigma}_j - \Sigma_{j+1}^0\| = v_T \|\Phi_j\| + o_p(v_T). \quad (\text{A.24})$$

When  $Mv_T^{-2} \leq \tau_{gl} - \tau_{g,l-1} \leq Bv_T^{-2} \log T$ , Property 2 together with (A.24) implies that

$$\bar{\ell}_{g,l}(\hat{\mathcal{K}}, \hat{\theta}) \leq -|O_p(M)|, \quad (\text{A.25})$$

for a sufficiently large  $M$ , while, for  $\tau_{gl} - \tau_{g,l-1} \leq Mv_T^{-2}$ , Property 4 with (A.24) implies

$$\bar{\ell}_{g,l}(\hat{\mathcal{K}}, \hat{\theta}) = O_p(M^{1/2}). \quad (\text{A.26})$$

Since  $\sup_{(\mathcal{K}, \theta) \in \ddot{\Xi}(B) \times \ddot{\Theta}(\epsilon)} \Delta(\mathcal{K}, \theta) = o(1)$ , Lemma A.4 with (A.23), (A.25) and (A.26) that

$$\sup_{(\mathcal{K}, \theta) \in \ddot{\Xi}(B) \setminus \ddot{\Xi}_M \times \ddot{\Theta}(\epsilon)} \ell_T(\mathcal{K}, \theta) < -|O_p(M)|,$$

for a sufficiently large  $M$ . This completes the proof of the part (a).

(b) From (a), there exists an  $M > 0$  such that  $\max_{1 \leq g \leq G} \max_{1 \leq j \leq m} |\hat{k}_{gj} - k_{gj}^0| \leq Mv_T^{-2}$  in probability. Thus it suffices to consider the case where either  $\tau_{gl} - \tau_{g,l-1} \leq Mv_T^{-2}$  or  $\tau_{gl} - \tau_{g,l-1} > \delta T$  for some  $\delta > 0$ . As in (A.23) and (A.26), we can show that  $\bar{\ell}_{g,l}(\hat{\mathcal{K}}, \hat{\theta})$  is bounded by  $|O_p(1)|$  for every  $(g, l) \in \{1, \dots, G+1\} \times \{1, \dots, 2(m+1)\}$ . If  $\sqrt{T} \|\hat{\beta}_{gj} - \beta_{gj}^0\| \geq M$  for some group and regime  $(g, j)$  and for some  $M > 0$ , then there is a corresponding sub-interval  $[\tau_{g,l-1} + 1, \tau_{gl}]$  with  $\tau_{gl} - \tau_{g,l-1} > \delta T$  and thus Property 3(a) shows that  $\bar{\ell}_{g,l}(\hat{\mathcal{K}}, \hat{\theta}) \leq -|O_p(M^2)|$  for a sufficiently large  $M$ . Thus, on the event that  $\max_{1 \leq g \leq G} \max_{1 \leq j \leq m+1} \|\hat{\beta}_{gj} - \beta_{gj}^0\| \geq MT^{-1/2}$  for a sufficiently large  $M$ , Lemma A.4 implies that the normalized log likelihood takes negative value in probability. The same result holds when  $\max_{1 \leq j \leq m+1} \|\hat{\Sigma}_j - \Sigma_j^0\| \geq MT^{-1/2}$  for a sufficiently large  $M$ . Hence the desired conclusion follows. ■

Having established convergence rates of estimates, we are now in a position to prove results about asymptotic independence of break date estimates and estimates of the basic parameters. In order to proceed, we need a notation to denote the likelihood based on observations in the interval  $[t_1, t_2] \subset [1, T]$  as  $L(t_1, t_2; \mathcal{K}, \theta) = \prod_{t=t_1}^{t_2} f(y_t | X_{tT}, \theta_{t,\mathcal{K}})$ . Then, using partitions  $\{[\tau_{l-1} + 1, \tau_l]\}_{l=1}^N$  of an interval  $[1, T]$  given  $\mathcal{K}$  and  $\mathcal{K}^0$ , we can express the

<sup>16</sup> To prove this, we use the inequality that  $\|a - b\| - \|b - c\| \leq \|a - c\| \leq \|a - b\| + \|b - c\|$ , for any elements  $a, b$  and  $c$  of some space with the norm  $\|\cdot\|$ , which is due to the triangle inequality.

normalized log likelihood as

$$\ell_T(\mathcal{K}, \theta) = \sum_{l=1}^N \left\{ \log L(\tau_{l-1} + 1, \tau_l; \mathcal{K}, \theta) - \log L(\tau_{l-1} + 1, \tau_l; \mathcal{K}^0, \theta^0) \right\}.$$

**Proof of Theorem 2.** Consider the case where  $(\mathcal{K}, \theta) \in \bar{\Xi}_M \times \bar{\Theta}_M$  for a sufficiently large  $M$  with the restriction  $R(\theta) = 0$ . From definition, we can write

$$\begin{aligned} \ell_T(\mathcal{K}, \theta) - \ell_T(\mathcal{K}^0, \theta) - \ell_T(\mathcal{K}, \theta^0) \\ = \sum_{l=1}^N \left\{ \log L(\tau_{l-1} + 1, \tau_l; \mathcal{K}, \theta) - \log L(\tau_{l-1} + 1, \tau_l; \mathcal{K}^0, \theta) \right\} \end{aligned} \quad (\text{A.27})$$

$$- \sum_{l=1}^N \left\{ \log L(\tau_{l-1} + 1, \tau_l; \mathcal{K}, \theta^0) - \log L(\tau_{l-1} + 1, \tau_l; \mathcal{K}^0, \theta^0) \right\}. \quad (\text{A.28})$$

If  $\tau_l - \tau_{l-1} > Mv_T^{-2}$ , then we have that  $\theta_{t,\mathcal{K}} = \theta_{t,\mathcal{K}^0}$  and  $\theta_{t,\mathcal{K}}^0 = \theta_{t,\mathcal{K}^0}^0$  for all  $\tau_{l-1} + 1 \leq t \leq \tau_l$ . Thus, it suffices to consider the quantities in (A.27) and (A.28) with the index  $l$  satisfying that  $\tau_l - \tau_{l-1} \leq Mv_T^{-2}$ . Property 4 with  $b_T = MT^{-1/2}$  implies that, uniformly in  $(\mathcal{K}, \theta) \in \bar{\Xi}_M \times \bar{\Theta}_M$ ,

$$\ell_T(\mathcal{K}, \theta) = \ell_T(\mathcal{K}, \theta^0) + \ell_T(\mathcal{K}^0, \theta) + O_p((\sqrt{T}v_T)^{-1}).$$

Hence we obtain the desired result. ■

To derive the limit distribution of the test, we shall present a technical lemma, which is a direct consequence of Lemma A.1(b). To this end, we introduce some notation. For  $j = 1, \dots, m$ , we define, for  $s < 0$ ,

$$V_{T,z\eta,j}^{(1)}(-s) := v_T \sum_{t=T_j^0+[sv_T^{-2}]}^{T_j^0} (z_t \otimes \eta_t) \quad \text{and} \quad V_{T,\eta\eta,j}^{(1)}(-s) := v_T \sum_{t=T_j^0+[sv_T^{-2}]}^{T_j^0} (\eta_t \eta_t' - I_n),$$

and, for  $s > 0$ ,

$$V_{T,z\eta,j}^{(2)}(s) := v_T \sum_{t=T_j^0}^{T_j^0+[sv_T^{-2}]} (z_t \otimes \eta_t) \quad \text{and} \quad V_{T,\eta\eta,j}^{(2)}(s) := v_T \sum_{t=T_j^0}^{T_j^0+[sv_T^{-2}]} (\eta_t \eta_t' - I_n).$$

**Lemma A.5.** Under Assumption B1-B2 with a sequence  $v_T$  defined in Assumption B4, we have, for  $j = 1, \dots, m$ ,

$$V_{T,z\eta,j}^{(1)}(\cdot) \Rightarrow \mathbb{V}_{z\eta,j}^{(1)}(\cdot) \quad \text{and} \quad V_{T,\eta\eta,j}^{(2)}(\cdot) \Rightarrow \mathbb{V}_{\eta\eta,j}^{(2)}(\cdot),$$

where the weak convergence is in the space  $D[0, \infty)^{nq}$  and the Brownian motions  $\mathbb{V}_{z\eta,j}^{(1)}(\cdot)$  and  $\mathbb{V}_{\eta\eta,j}^{(2)}(\cdot)$  are defined in the main text. Furthermore, for  $j = 1, \dots, m$ ,

$$V_{T,\eta\eta,j}^{(1)}(\cdot) \Rightarrow \mathbb{V}_{\eta\eta,j}^{(1)}(\cdot) \quad \text{and} \quad V_{T,z\eta,j}^{(2)}(\cdot) \Rightarrow \mathbb{V}_{z\eta,j}^{(2)}(\cdot),$$

where the weak convergence is in the space  $D[0, \infty)^{n^2}$  and the  $n \times n$  matrices  $\mathbb{V}_{\eta\eta,j}^{(1)}(\cdot)$  and  $\mathbb{V}_{z\eta,j}^{(2)}(\cdot)$  are Brownian motion defined in the main text.<sup>17</sup>

---

<sup>17</sup> Note that  $\mathbb{V}_{z\eta,j}^{(1)}(\cdot)$  and  $\mathbb{V}_{\eta\eta,j}^{(1)}(\cdot)$  are not necessarily independent unless  $E[\eta_{tk} \otimes \eta_{tl} \otimes \eta_{th}] = 0$  for all  $k, l, h$  and for every  $t$ . The argument can apply for  $\mathbb{V}_{z\eta,j}^{(2)}(\cdot)$  and  $\mathbb{V}_{\eta\eta,j}^{(2)}(\cdot)$ .

**Proof of Lemma 1.** Consider the regime  $j \in \{1, \dots, m\}$ . For  $s \in \mathbb{R}$  and for  $\underline{T}_j^0(s) \leq t \leq \bar{T}_j^0(s)$ , observe that

$$(\Sigma_{t,\mathcal{K}}^0)^{-1} = \begin{cases} (\Sigma_{j+1}^0)^{-1}, & \text{if } k_{Gj} \leq \underline{T}_j^0(s) \\ (\Sigma_{j+1}^0)^{-1} - \mathbb{1}_{\{T_j^0 < t \leq k_{Gj}\}} \{(\Sigma_{j+1}^0)^{-1} - (\Sigma_j^0)^{-1}\}, & \text{if } T_j^0 < k_{Gj} \leq T_j^0(s) \\ (\Sigma_j^0)^{-1} + \mathbb{1}_{\{k_{Gj} < t \leq T_j^0\}} \{(\Sigma_{j+1}^0)^{-1} - (\Sigma_j^0)^{-1}\}, & \text{if } T_j^0(s) < k_{Gj} \leq T_j^0 \\ (\Sigma_j^0)^{-1}, & \text{if } \bar{T}_j^0(s) \leq k_{Gj}, \end{cases}$$

which yields

$$(\Sigma_{t,\mathcal{K}}^0)^{-1} = (\Sigma_{j+1}^0)^{-1} - \text{sgn}(s_{Gj}) \mathbb{1}_{\{|s_{Gj}| \leq |s|\}} \{(\Sigma_{j+1}^0)^{-1} - (\Sigma_j^0)^{-1}\}.$$

Let  $D_{T,j}(s) := v_T^2 \sum_{t=\underline{T}_j^0(s)+1}^{\bar{T}_j^0(s)} x_{tT} x'_{tT}$ . Since  $X_{tT}(\Sigma_{t,\mathcal{K}}^0)^{-1} X'_{tT} = S' x_{tT} x'_{tT} \otimes (\Sigma_{t,\mathcal{K}}^0)^{-1} S$ , we have

$$\begin{aligned} B_{T,j}(s, s_{Gj}) &= S' D_{T,j}(s) \otimes (\Sigma_{j+1}^0)^{-1} S \\ &\quad - \text{sgn}(s_{Gj}) \mathbb{1}_{\{|s_{Gj}| \leq |s|\}} S' D_{T,j}(s_{Gj}) \otimes \{(\Sigma_{j+1}^0)^{-1} - (\Sigma_j^0)^{-1}\} S, \end{aligned}$$

for every  $\underline{T}_j^0(s) \leq t \leq \bar{T}_j^0(s)$ . We can show that, for  $\underline{T}_j^0(s) \leq t \leq \bar{T}_j^0(s)$ ,

$$\varphi(t/T) = \varphi(\lambda_j^0) + O((\sqrt{T}v_T)^{-2}) \quad \text{and} \quad w_t = w_{T_j^0} + O((\sqrt{T}v_T)^{-2}), \quad (\text{A.29})$$

uniformly in  $s \in \mathbb{R}$ .<sup>18</sup> Also, under Assumption B1, we can show that, uniformly in  $s \in \mathbb{R}$ ,

$$v_T^2 \sum_{t=\underline{T}_j^0(s)+1}^{\bar{T}_j^0(s)} z_t = |s| \mu_{z,j+1} \mathbb{1}_{\{0 < s\}} + o_p(1) \quad \text{and} \quad v_T^2 \sum_{t=\underline{T}_j^0(s)+1}^{\bar{T}_j^0(s)} z_t z'_t = |s| Q_{zz,j+1} \mathbb{1}_{\{0 < s\}} + o_p(1).$$

It follows that, uniformly in  $s \in \mathbb{R}$ ,

$$D_{T,j}(s) = |s| \begin{pmatrix} Q_{zz,j+1} \mathbb{1}_{\{0 < s\}} & \mu_{z,j+1} \mathbb{1}_{\{0 < s\}} \varphi(\lambda_j^0)' & \mu_{z,j+1} \mathbb{1}_{\{0 < s\}} T^{-1/2} w'_{T_j^0} \\ \varphi(\lambda_j^0) \mu'_{z,j+1} \mathbb{1}_{\{0 < s\}} & \varphi(\lambda_j^0) \varphi(\lambda_j^0)' & \varphi(\lambda_j^0) T^{-1/2} w'_{T_j^0} \\ T^{-1/2} w_{T_j^0} \mu'_{z,j+1} \mathbb{1}_{\{0 < s\}} & T^{-1/2} w_{T_j^0} \varphi(\lambda_j^0)' & (T^{-1/2} w_{T_j^0})(T^{-1/2} w_{T_j^0})' \end{pmatrix} + o_p(1).$$

Also, we have that  $X_{tT}(\Sigma_{t,\mathcal{K}}^0)^{-1} u_t = S'(I \otimes (\Sigma_{t,\mathcal{K}}^0)^{-1})(x_{tT} \otimes u_t)$  and  $u_t = (\Sigma_{j+1}^0)^{1/2} \eta_t$ . Thus, we have, for  $\underline{T}_j^0(s) \leq t \leq \bar{T}_j^0(s)$ ,

$$\begin{aligned} W_{T,j}(s, s_{Gj}) &= S'(I_q \otimes (\Sigma_{j+1}^0)^{-1}) V_{T,j}(s) \\ &\quad - \text{sgn}(s_{Gj}) \mathbb{1}_{\{|s_{Gj}| \leq |s|\}} S'(I_q \otimes \{(\Sigma_{j+1}^0)^{-1} - (\Sigma_j^0)^{-1}\}) V_{T,j}(s_{Gj}), \end{aligned}$$

where  $V_{T,j}(s) := (I_q \otimes (\Sigma_{j+1}^0)^{1/2}) v_T \sum_{t=\underline{T}_j^0(s)+1}^{\bar{T}_j^0(s)} (x_{tT} \otimes \eta_t)$ . Using (A.29), we can show

$$v_T \sum_{t=\underline{T}_j^0(s)+1}^{\bar{T}_j^0(s)} (x_{tT} \otimes \eta_t) = \left( v_T \sum_{t=\underline{T}_j^0(s)+1}^{\bar{T}_j^0(s)} (z_t \otimes \eta_t)', \left( \varphi(\lambda_j^0)', T^{-1/2} w'_{T_j^0} \right) \otimes v_T \sum_{t=\underline{T}_j^0(s)+1}^{\bar{T}_j^0(s)} \eta_t' \right)' + o_p(1),$$

uniformly in  $s \in \mathbb{R}$ . Hence, Lemma A.5 with the continuous mapping theorem yields that  $\{B_{T,j}(\cdot), V_{T,j}(\cdot)\}_{j=1}^m \Rightarrow \{\mathbb{B}_j(\cdot), \mathbb{V}_j(\cdot)\}_{j=1}^m$ .  $\blacksquare$

<sup>18</sup> We have that  $a^r - b^r = (a - b) \sum_{l=0}^{r-1} a^{r-1-l} b^l$  for  $a, b \in \mathbb{R}$  and for an integer  $r \geq 2$ . It follows that  $|(t/T)^r - (T_j^0/T)^r| \leq C|(t - T_j^0)/T|$ .

**Proof of Theorem 3.** Consider the regime  $j \in \{1, \dots, m\}$ . We can write, for  $1 \leq t \leq T$ ,

$$\begin{aligned} \log f(y_t | X_{tT}, \theta_{t,\mathcal{K}}^0) &= -\frac{1}{2} \left\{ \log(2\pi)^n + \log |\Sigma_{t,\mathcal{K}}^0| + \|(\Sigma_{t,\mathcal{K}}^0)^{-1/2} u_t\|^2 \right. \\ &\quad \left. - 2(\Delta\beta_{t,\mathcal{K}}^0)' X_{tT}(\Sigma_{t,\mathcal{K}}^0) u_t + \|(\Sigma_{t,\mathcal{K}}^0)^{-1/2} X_{tT}' \Delta\beta_{t,\mathcal{K}}^0\|^2 \right\}. \end{aligned}$$

Define  $\underline{k}_{Gj} := \min\{k_{Gj}, T_j^0\}$  and  $\bar{k}_{Gj} := \max\{k_{Gj}, T_j^0\}$ . We have that

$$\sum_{\underline{k}_j+1}^{\bar{k}_j} \left\{ \log f(y_t | X_{tT}, \theta_{t,\mathcal{K}}^0) - \log f(y_t | X_{tT}, \theta_{t,\mathcal{T}^0}^0) \right\} = \ell_j^{(1)} + \ell_j^{(2)},$$

where

$$\begin{aligned} \ell_j^{(1)} &:= \frac{1}{2} \sum_{t=\underline{k}_{Gj}+1}^{\bar{k}_{Gj}} \left\{ \log |\Sigma_{t,\mathcal{T}^0}^0 (\Sigma_{t,\mathcal{K}}^0)^{-1}| + \text{tr} \left( \{(\Sigma_{t,\mathcal{T}^0}^0)^{-1} - (\Sigma_{t,\mathcal{K}}^0)^{-1}\} u_t u_t' \right) \right\}, \\ \ell_j^{(2)} &:= \frac{1}{2} \sum_{t=\underline{k}_j+1}^{\bar{k}_j} \left\{ 2(\Delta\beta_{t,\mathcal{K}}^0)' X_{tT}(\Sigma_{t,\mathcal{T}^0}^0)^{-1} u_t - \|(\Sigma_{t,\mathcal{T}^0}^0)^{-1/2} X_{tT}' \Delta\beta_{t,\mathcal{K}}^0\|^2 \right\}. \end{aligned}$$

We consider the term  $\ell_j^{(1)}$ . We can write  $\Sigma_{t,\mathcal{T}^0}^0 (\Sigma_{t,\mathcal{K}}^0)^{-1} = I_n + (\Sigma_{t,\mathcal{T}^0}^0 - \Sigma_{t,\mathcal{K}}^0) (\Sigma_{t,\mathcal{K}}^0)^{-1}$  and  $\Sigma_{t,\mathcal{K}}^0 - \Sigma_{t,\mathcal{T}^0}^0 = v_T \Phi_{t,\mathcal{K}}$ , where  $\Phi_{t,\mathcal{K}} = \Phi_j$  if  $k_{Gj} < t \leq T_j^0$  and  $\Phi_{t,\mathcal{K}} = -\Phi_j$  if  $T_j^0 < t \leq k_{Gj}$ . Thus, an application of the Taylor series expansion yields that, for  $\underline{k}_{Gj} \leq t \leq \bar{k}_{Gj}$ ,

$$\log |\Sigma_{t,\mathcal{T}^0}^0 (\Sigma_{t,\mathcal{K}}^0)^{-1}| = \text{tr} \left( -v_T \Phi_{t,\mathcal{K}} (\Sigma_{t,\mathcal{K}}^0)^{-1} \right) + \frac{1}{2} \text{tr} \left( \{v_T \Phi_{t,\mathcal{K}} (\Sigma_{t,\mathcal{K}}^0)^{-1}\}^2 \right) + O_p(v_T^3). \quad (\text{A.30})$$

Also we can write  $(\Sigma_{t,\mathcal{T}^0}^0)^{-1} - (\Sigma_{t,\mathcal{K}}^0)^{-1} = (\Sigma_{t,\mathcal{T}^0}^0)^{-1} (\Sigma_{t,\mathcal{K}}^0 - \Sigma_{t,\mathcal{T}^0}^0) (\Sigma_{t,\mathcal{K}}^0)^{-1}$  and  $u_t = (\Sigma_{t,\mathcal{T}^0}^0)^{1/2} \eta_t$ , which implies, for  $\underline{k}_{Gj} \leq t \leq \bar{k}_{Gj}$ ,

$$\text{tr} \left( \{(\Sigma_{t,\mathcal{T}^0}^0)^{-1} - (\Sigma_{t,\mathcal{K}}^0)^{-1}\} u_t u_t' \right) = \text{tr} \left( (\Sigma_{t,\mathcal{T}^0}^0)^{-1/2} v_T \Phi_{t,\mathcal{K}} (\Sigma_{t,\mathcal{K}}^0)^{-1} (\Sigma_{t,\mathcal{T}^0}^0)^{1/2} \eta_t \eta_t' \right). \quad (\text{A.31})$$

Given  $\underline{k}_{Gj} \leq t \leq \bar{k}_{Gj}$  with  $k_{Gj} = T_j^0 + [s_{Gj} v_T^{-2}]$ , we have that  $(\Phi_{t,\mathcal{K}}, \Sigma_{t,\mathcal{T}^0}^0, \Sigma_{t,\mathcal{K}}^0)$  becomes  $(\Phi_j, \Sigma_j^0, \Sigma_{j+1}^0)$  for  $s_{Gj} \leq 0$  and  $(-\Phi_j, \Sigma_{j+1}^0, \Sigma_j^0)$  for  $s_{Gj} > 0$ . This implies that  $(\Sigma_{t,\mathcal{T}^0}^0)^{-1/2} \Phi_{t,\mathcal{K}} (\Sigma_{t,\mathcal{K}}^0)^{-1} (\Sigma_{t,\mathcal{T}^0}^0)^{1/2} = \Pi_j(s_{Gj})$ , which together with (A.30) and (A.31) yields

$$\ell_j^{(1)} = \frac{1}{2} \text{tr} \left( \Pi_j(s_{Gj}) V_{T,\eta\eta,j}(s_G) \right) + \frac{|s_{Gj}|}{4} \text{tr} \left( \{\Pi_j(s_{Gj})\}^2 \right) + o_p(1),$$

where  $V_{T,\eta\eta,j}(s_G) := v_T \sum_{t=\underline{k}_{Gj}+1}^{\bar{k}_{Gj}} (\eta_t \eta_t' - I_n)$ .

We consider the term  $\ell_j^{(2)}$ . Define  $\Delta\beta_{g,t,\mathcal{K}}^0 := \sum_{i \in \mathcal{G}_g} e_i \circ (\beta_{t,\mathcal{K}}^0 - \beta_{t,\mathcal{T}^0}^0)$ . Then  $\Delta\beta_{t,\mathcal{K}}^0 = \sum_{g=1}^G \Delta\beta_{g,t,\mathcal{K}}^0$  and we have

$$\ell_j^{(2)} = \sum_{t=\underline{k}_j+1}^{\bar{k}_j} \left( \sum_{g=1}^G (\Delta\beta_{g,t,\mathcal{K}}^0)' X_{tT}(\Sigma_{t,\mathcal{K}}^0)^{-1} u_t - \frac{1}{2} \sum_{g=1}^G \sum_{l=1}^G (\Delta\beta_{g,t,\mathcal{K}}^0)' X_{tT}(\Sigma_{t,\mathcal{K}}^0)^{-1} X_{tT}' \Delta\beta_{l,t,\mathcal{K}}^0 \right).$$

For the group  $g \in \{1, \dots, G\}$ , we have that  $\Delta\beta_{g,t,\mathcal{K}}^0 = \beta_{g,j+1}^0 - \beta_{gj}^0$  for  $k_{gj} < t \leq T_j^0$  and that

$\Delta\beta_{g,t,\mathcal{K}}^0 = -(\beta_{g,j+1}^0 - \beta_{gj}^0)$  for  $T_j^0 < t \leq k_{gj}$ . It follows that

$$\sum_{t=\bar{k}_j+1}^{\bar{k}_j} (\Delta\beta_{g,t,\mathcal{K}}^0)' X_{tT} (\Sigma_{t,\mathcal{K}}^0)^{-1} u_t = -\text{sgn}(s_{gj}) \delta'_{gj} W_{T,j}(s_{gj}, s_{Gj}).$$

Similarly, for groups  $g, l \in \{1, \dots, G\}$ , we have that

$$\begin{aligned} \sum_{t=\bar{k}_j+1}^{\bar{k}_j} (\Delta\beta_{g,t,\mathcal{K}}^0)' X_{tT} (\Sigma_{t,\mathcal{K}}^0)^{-1} X'_{tT} \Delta\beta_{l,t,\mathcal{K}}^0 &= \mathbb{1}_{\{k_{gj} \vee k_{lj} \leq T_j^0\}} \delta'_{gj} B_{T,j}(s_{gj} \vee s_{lj}, s_{Gj}) \delta_{lj} \\ &\quad + \mathbb{1}_{\{T_j^0 < k_{gj} \wedge k_{lj}\}} \delta'_{gj} B_{T,j}(s_{gj} \wedge s_{lj}, s_{Gj}) \delta_{lj}. \end{aligned}$$

Thus,

$$\begin{aligned} \ell_j^{(1)} &= -\text{sgn}(s_{gj}) \delta'_{gj} W_{T,j}(s_{gj}, s_{Gj}) \\ &\quad - \frac{1}{2} \sum_{g=1}^G \sum_{l=1}^G \delta'_{gj} \left\{ \mathbb{1}_{\{s_{gj} \vee s_{lg} \leq 0\}} B_{T,j}(s_{gj} \vee s_{lg}, s_{Gj}) + \mathbb{1}_{0 < \{s_{gj} \wedge s_{lg}\}} B_{T,j}(s_{gj} \wedge s_{lg}, s_{Gj}) \right\} \delta_{lj}. \end{aligned}$$

Applying Lemma 1 and the continuous mapping theorem, we obtain the desired result.  $\blacksquare$

**Proof of Theorem 4.** Under both alternatives  $H_1$  and  $H_{1T}$ , the convergence rates of Theorem 1 can apply for estimates  $\hat{\theta}$  and  $\hat{\mathcal{K}}$  obtained from the model under the alternative. Thus, given collections of break dates  $\hat{\mathcal{K}}$  and  $\mathcal{K}^0$ , sub-intervals  $\{[\tau_{g,l-1} + 1, \tau_{gl}]\}_{l=1}^{N_g}$  for each group  $g$  satisfy either  $\tau_{gl} - \tau_{g,l-1} \geq \nu T$  or  $\tau_{gl} - \tau_{g,l-1} \leq M v_T^{-2}$  for some  $M > 0$ . If  $\tau_{gl} - \tau_{g,l-1} \geq \nu T$ , then the argument used to prove Property 3(b) with  $\sqrt{T}$ -consistent estimate  $\hat{\theta}$  shows that  $\bar{\ell}_{g,l}(\hat{\mathcal{K}}, \hat{\theta}) = O_p(1)$ , while If  $\tau_{gl} - \tau_{gl} \leq M v_T^{-2}$ , then the argument to obtain (A.26) shows that  $\bar{\ell}_{g,l}(\hat{\mathcal{K}}, \hat{\theta}) = O_p(1)$ . Also, Theorem 1(b) implies that  $\Delta(\hat{\mathcal{K}}, \hat{\theta}) = o_p(1)$ . It follows from Lemma A.4 that

$$\ell_T(\hat{\mathcal{K}}, \hat{\theta}) = O_p(1). \quad (\text{A.32})$$

It remains to consider the normalized likelihood  $\ell_T(\tilde{\mathcal{T}}, \tilde{\theta})$  obtained under the null  $H_0$ .

(a) Let  $\delta \in (0, 1)$  be fixed. If  $\max_{1 \leq j \leq m} \max_{1 \leq g_1, g_2 \leq G} |k_{g_1 j}^0 - k_{g_2 j}^0| \geq \delta T$ , then we have  $\max_{1 \leq j \leq m} \max_{1 \leq g \leq G} |\tilde{T}_j - k_{gj}^0| \geq \delta T/2$ . Applying a similar argument used in Proposition A.1, we can show that Property 1 and Property 2 with  $m_T = \delta T/2$  imply that

$$\ell_T(\tilde{\mathcal{T}}, \tilde{\theta}) \leq -|O_p(T v_T^2)|. \quad (\text{A.33})$$

It follows from (A.32) and (A.33) that  $CB_T = \ell_T(\hat{\mathcal{K}}, \hat{\theta}) - \ell_T(\tilde{\mathcal{T}}, \tilde{\theta}) \geq |O_p(T v_T^2)|$ . Since the critical value  $c_\alpha^*$  is a finite value, we obtain the desired result.

(b) If  $\max_{1 \leq j \leq m} \max_{1 \leq g_1, g_2 \leq G} |k_{g_1 j}^0 - k_{g_2 j}^0| \geq M v_T^{-2}$  for some constant  $M > 0$ , then we have  $\max_{1 \leq j \leq m} \max_{1 \leq g \leq G} |\tilde{T}_j - k_{gj}^0| \geq M v_T^{-2}/2$ . When  $\max_{1 \leq j \leq m} \max_{1 \leq g \leq G} |\tilde{T}_j - k_{gj}^0| \geq D v_T^{-2} \log T$  for a sufficiently large  $D$ , it is shown that  $\ell_T(\tilde{\mathcal{T}}, \tilde{\theta}) \leq -|O_p(M)|$  in the proof of Proposition A.1. When  $M v_T^{-2} \leq \max_{1 \leq j \leq m} \max_{1 \leq g \leq G} |\tilde{T}_j - k_{gj}^0| \leq D v_T^{-2} \log T$ , it follows from the proof of Theorem 1(a) that  $\ell_T(\tilde{\mathcal{T}}, \tilde{\theta}) \leq -|O_p(M)|$  for a sufficiently large  $M > 0$ . Thus, there is some  $M > 0$  such that  $CB_T \geq |O_p(M)|$  and the proof is completed.  $\blacksquare$

Table 1: Empirical Rejection Frequencies under the null hypothesis

		AR Coefficient								
		$\alpha = 0.0$			$\alpha = 0.4$			$\alpha = 0.8$		
Break Size		Nominal Size			Nominal Size			Nominal Size		
$\delta_1$	$\delta_2$	10%	5%	1%	10%	5%	1%	10%	5%	1%
0.50	0.50	0.064	0.036	0.004	0.086	0.050	0.004	0.162	0.104	0.032
	0.75	0.070	0.036	0.004	0.094	0.054	0.006	0.158	0.088	0.032
	1.00	0.084	0.036	0.004	0.106	0.060	0.010	0.170	0.098	0.038
	1.25	0.086	0.044	0.004	0.108	0.058	0.014	0.182	0.104	0.040
	1.50	0.096	0.050	0.006	0.120	0.056	0.010	0.186	0.108	0.036
0.75	0.75	0.084	0.032	0.004	0.112	0.046	0.004	0.158	0.086	0.030
	1.00	0.088	0.040	0.004	0.108	0.050	0.010	0.154	0.082	0.030
	1.25	0.086	0.050	0.006	0.104	0.060	0.006	0.156	0.088	0.028
	1.50	0.090	0.052	0.006	0.118	0.058	0.010	0.166	0.090	0.028
1.00	1.00	0.090	0.044	0.008	0.104	0.060	0.012	0.150	0.078	0.022
	1.25	0.086	0.050	0.010	0.090	0.060	0.010	0.140	0.072	0.026
	1.50	0.092	0.050	0.012	0.096	0.056	0.012	0.152	0.070	0.026
1.25	1.25	0.080	0.044	0.008	0.084	0.052	0.012	0.118	0.058	0.018
	1.50	0.074	0.042	0.010	0.080	0.044	0.010	0.112	0.056	0.018
1.50	1.50	0.074	0.038	0.010	0.088	0.040	0.010	0.106	0.048	0.018

*Notes:* The data generating process is the bivariate system:

$$y_{1t} = 1 + \delta_1 \mathbb{1}_{\{k_1 < t\}} + \alpha y_{1,t-1} + u_{1t} \quad (\text{EQ1})$$

$$y_{2t} = 1 + \delta_2 \mathbb{1}_{\{k_2 < t\}} + \alpha y_{2,t-1} + u_{2t}, \quad (\text{EQ2})$$

for  $t = 1, \dots, T$ , where  $(u_{1t}, u_{2t})' \sim i.i.d.N(0, I_2)$  and  $\delta_i$  is the break size for the  $i^{th}$  equation for  $i = 1, 2$ . We set the sample size  $T = 100$ , the break date  $k_1 = k_2 = 50$  and the trimming value  $\nu = 0.15$ .



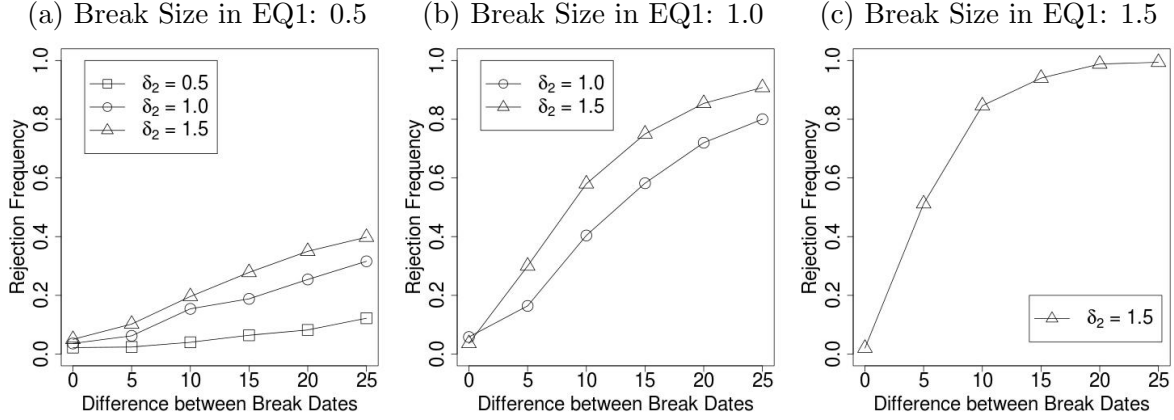
Table 2: Structural breaks in the U.S. disaggregated inflation series

Replication of the results in Clark (2006)			
	OLS without break		
	Durables	Nondurables	Service
	Persistence	0.921	0.878
	OLS with common break		
	Durables	Nondurables	Service
	Persistence	0.800	0.367
Break Date (Known)	93:Q1		
Evidence from SUR system			
	SUR with common breaks ( $k_1 = k_2 = k_3$ )		
	Durables	Nondurables	Service
	Persistence	0.805	0.356
Break Date	92:Q1		
Test for Common Break			
Null Hypothesis	LR test	Critical value (5%)	
$H_0 : k_1 = k_2 = k_3$	9.051	3.858	
$H_0 : k_1 = k_2$	9.735	1.871	
$H_0 : k_1 = k_3$	7.684	3.003	
$H_0 : k_2 = k_3$	0.749	2.762	
	SUR with common break ( $k_2 = k_3$ )		
	Durables	Nondurables	Service
	Persistence	0.324	0.406
Break Date	95:Q1	92:Q1	
95% C.I.	[94:Q2, 95:Q4]	[91:Q3, 92:Q3]	

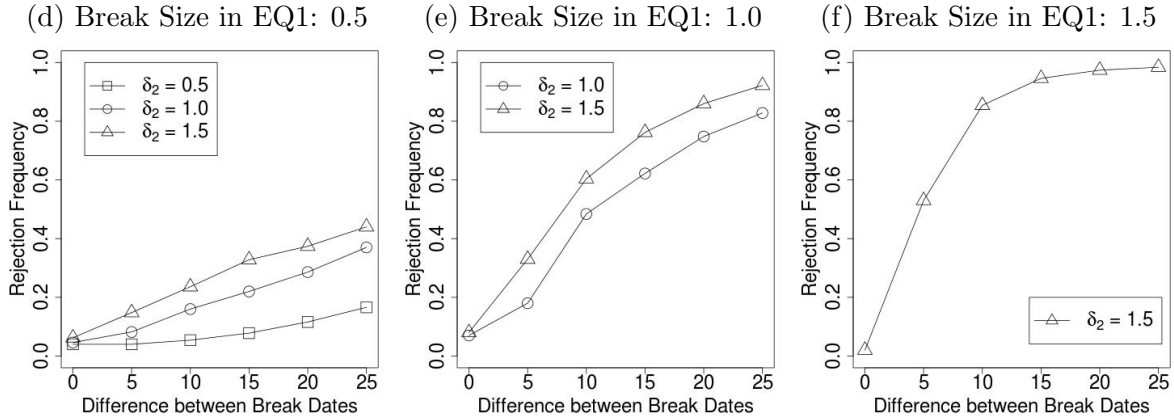
*Notes:* The sample period is 1984 to 2002. The estimated model is the AR model with the intercept and the AR lag length selected by the AIC is 4, 5 or 3 for durables, nondurables or service, respectively. *Persistence* is measured by the sum of AR coefficients. The critical values at the 5% significance level are obtained through simulation with 3,000 repetitions. C.I. denotes the 95% confidence interval of the break date.

Figure 1: Finite-sample power of the test

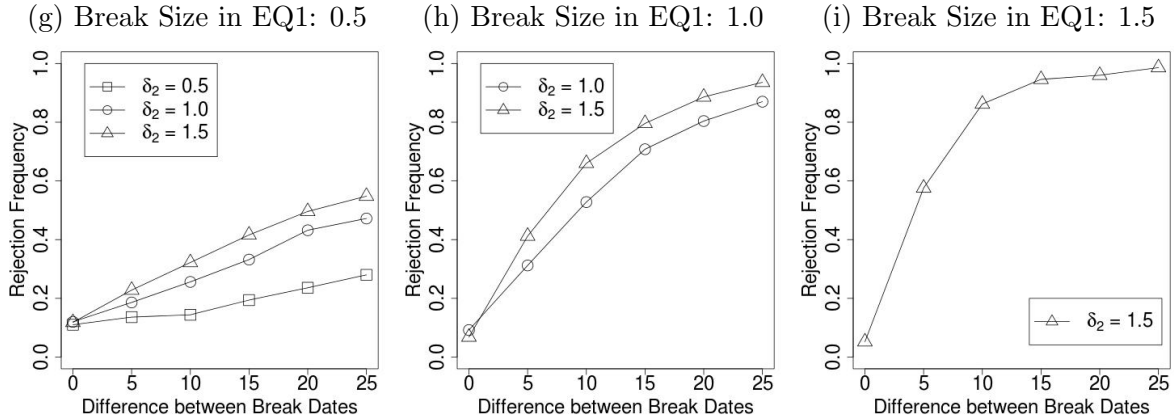
Panel A: AR Coefficient = 0.00



Panel B: AR Coefficient = 0.40



Panel C: AR Coefficient = 0.80



*Notes:* The data generating process is the bivariate system as in (EQ1) and (EQ2) of Table 1. The number of observations  $T$  is set to 100. The break date  $k_1$  in (EQ1) is kept fixed at  $k_1 = 35$ , while the break date  $k_2$  in (EQ2) changes from 30 to 55. The horizontal axis shows the difference between break dates:  $k_2 - k_1$ . The AR coefficient  $\alpha$  is set to 0.0, 0.4 and 0.8 for Panel A, B and C, respectively. The break size  $\delta_1$  in (EQ1) changes across panel (a)-(c), (d)-(f) and (g)-(i), while the break size  $\delta_2$  in (EQ2) changes within each panel. We use 0.5, 1.0 and 1.5 as magnitude of the break size.